

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
21 February 2002 (21.02.2002)

PCT

(10) International Publication Number
WO 02/14482 A2

(51) International Patent Classification⁷: C12N 9/00

(21) International Application Number: PCT/GB01/03642

(22) International Filing Date: 14 August 2001 (14.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0019986.9 14 August 2000 (14.08.2000) GB

(71) Applicant (for all designated States except US): BIOTICA
TECHNOLOGY LIMITED [GB/GB]; 181A Huntingdon
Road, Cambridge, Cambridgeshire CB3 0DJ (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PETKOVIC,

Hrvoje [SI/GB]; 89 Argyle Street, Cambridge, Cam-
bridgeshire CB1 3LS (GB). KENDREW, Steven, Gary
[GB/GB]; 6 Cowper Road, Cambridge, Cambridgeshire
CB1 3SN (GB). LEADLAY, Peter, Francis [GB/GB];
6 Westberry Court, Pinehurst, Grange Road, Cambridge,
Cambridgeshire CB3 9BG (GB).

(74) Agents: STUART, Ian et al.; Mewburn Ellis, York House,
23 Kingsway, London, Greater London WC2B 6HP (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZW.

[Continued on next page]

(54) Title: POLYKETIDES AND THEIR SYNTHESIS



(57) Abstract: Biosyntheses of compounds
whereof at least portions are polyketides
produced by means of polyketide synthase
(PKS) enzyme complexes are carried out
after specific alterations have been made
within the acyltransferase (AT) domains
of the PKSs. Particular motifs in or near
the substrate binding pocket are disclosed,
such that alterations therein affect substrate
specificity.

WO 02/14482 A2



(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Polyketides and Their Synthesis

Technical Field

The present invention relates to processes and materials (including enzyme systems, nucleic acids, vectors and cultures) which can be used to influence the selection of acylthioester units for the synthesis of polyketides, and to the resulting polyketides, which may be novel. It is particularly concerned with macrolides, polyethers or polyenes and their preparation making use of recombinant synthesis.

In preferred types of embodiment, polyketide biosynthetic genes or portions of them, which may be derived from different polyketide biosynthetic gene clusters, are manipulated to allow the production of specific polyketides, such as 12-, 14- and 16-membered macrolides, of predicted structure. The invention is particularly concerned with the modification of an Acyl CoA:ACP transferase (AT) function, generally by modifying genetic material encoding it in order to prepare polyketides with a predetermined ketide unit, e.g. incorporating (a) a malonate extender unit; or (b) a methylmalonate extender unit; or (c) an ethylmalonate extender unit; or (d) a further type of extender unit; or (e) an acetate and/or malonate starter unit; or (f) a

propionate and/or methylmalonate starter unit; or (g) a butyrate and/or ethylmalonate starter unit; or (h) a further type of starter unit. Of course the invention can be used to influence more than one ketide unit of a polyketide. The method enables one to minimise alteration to the protein structure of the polyketide synthase.

Polyketides are a large and structurally diverse class of natural products that includes many compounds possessing antibiotic or other pharmacological properties, such as erythromycin, tetracyclines, rapamycin, avermectin, monensin, epothilone and FK506. In particular, polyketides are abundantly produced by *Streptomyces* and related actinomycete bacteria. They are synthesised by the repeated stepwise condensation of acylthioesters in a manner analogous to that of fatty acid biosynthesis. The structural diversity found among natural polyketides arises in part from the selection of (usually) acetate (malonyl-CoA) or propionate (methylmalonyl-CoA) as "starter" or "extender" units (although one of a variety of other types of unit may occasionally be selected); as well as from the differing degree of processing of the β -keto group formed after each condensation. Examples of processing steps include reduction to β -hydroxyacyl-, reduction followed by

dehydration to 2-enoyl-, and complete reduction to the saturated acylthioester. The stereochemical outcome of these processing steps is also specified for each cycle of chain extension. Methylation at the α -carbon or β -hydroxy is also sometimes observed.

The biosynthesis of polyketides is performed by a group of chain-forming enzymes known as polyketide synthases. Two broad classes of polyketide synthase (PKS) have been described in actinomycetes. One class, named Type I PKSs, represented by the PKSs for the macrolides erythromycin, oleandomycin, avermectin, and rapamycin and by the PKS for the polyether monensin, consists of a different set or "module" of enzymes for each cycle of polyketide chain extension. For an example see Figure 1 (Cortés, J. et al. *Nature* (1990) 348:176-178; Donadio, S. et al. *Science* (1991) 2523:675-679; Swan, D.G. et al. *Mol. Gen. Genet.* (1994) 242:358-362; MacNeil, D. J. et al. *Gene* (1992) 115:119-125; Schwecke, T. et al. *Proc. Natl. Acad. Sci. USA* (1995) 92:7839-7843; also Patent application WO98/01546). The genes encoding numerous Type I PKSs have been sequenced and these sequences disclosed in publicly available DNA and protein sequence databases including Genbank, Swissprot and EMBL. For example, the sequences are available for the PKSs

governing the synthesis of erythromycin (Cortes, J. et al. Nature (1990) 348:176-178); accession number X62569, Donadio, S. et al. Science (1991) 252:675-679; accession number M63677); rapamycin (Schwecke, T. et al. Proc. Natl. Acad. Sci. (1995) 92:7839-7843; accession number X86780); rifamycin (August, P. et al. Chem. Biol. (1998) 5:69-79; accession number AF040570) and tylosin (Eli Lilly, accession number U78289), among many others.

The term "polyketide synthase" (PKS) as used herein refers to a complex of enzyme activities responsible for the biosynthesis of polyketides. These enzyme activities include β -ketoacyl ACP synthase (KS), acyltransferase (AT), acyl carrier protein (ACP), β -ketoreductase (KR), dehydratase (DH), enoylreductase (ER) and thioesterase (TE) but are not limited to these activities. Each of these activities lies on a separate protein or polypeptide fragment responsible for this activity. Such a fragment is termed a "domain". The terms "motif" or "signature sequence" used herein refer to a small stretch of amino acids (usually less than 10 amino acids) within a domain responsible (at least in part) for one aspect of the catalytic function, for example, choice of substrate.

The term "extension module" as used herein refers to the set of contiguous domains, from a β -ketoacyl-ACP synthase

("KS") domain to the next acyl carrier protein ("ACP") domain, which accomplishes one cycle of polyketide chain extension; this may or may not include domains responsible for the reductive processing of the polyketide chain. The term "loading module" is used to refer to any group of contiguous domains that accomplishes the loading of the starter unit onto the PKS and thus renders it available to the KS domain of a specific extension module.

10

Background Art

Several approaches to altering the nature of the polyketide product of a PKS by genetic engineering have been proposed: see particularly WO 93/13663 and WO 98/01571. The length of polyketide formed has been altered, in the case of erythromycin biosynthesis, by specific relocation using genetic engineering of the enzymatic domain of the erythromycin-producing PKS that contains the chain-releasing thioesterase/cyclase activity (Cortés, J. et al. Science (1995) 268:1487-1489; Kao, C.M. et al. J. Am. Chem. Soc. (1995) 117:9105-9106).

In-frame deletion of the DNA encoding part of the ketoreductase domain in module 5 of the erythromycin-producing PKS (also known as 6-deoxyerythronolide B synthase, DEBS) has been shown to lead to the formation

25

of erythromycin analogues 5,6-dideoxy-3- α -mycarosyl-5-oxoerythronolide B, 5,6-dideoxy-5-oxoerythronolide B and 5,6-dideoxy, 6 β -epoxy-5-oxoerythronolide B (Donadio, S. et al. Science (1991) 252:675-679). Likewise, alteration
5 of active site residues in the enoylreductase domain of module 4 in DEBS, by genetic engineering of the corresponding PKS-encoding DNA and its introduction into *Saccharopolyspora erythraea*, led to the production of 6,7-anhydroerythromycin C (Donadio, S. et al. Proc Natl.
10 Acad. Sci. USA (1993) 90:7119-7123).

Patent application WO 00/01827 describes further methods of manipulating a PKS to change the oxidation state of the β -carbon. Substituting the reductive domain of module 2 of the erythromycin-producing PKS with
15 domains derived from rapamycin PKS modules 10 and 13 led to the formation of C10-C11 olefin-erythromycin A and C10-C11 dihydroerythromycin A respectively.

The second class of PKS, named Type II PKSS, is represented by the synthases for aromatic compounds.
20 Type II PKSS contain only a single set of enzymatic activities for chain extension and these are re-used as appropriate in successive cycles (Bibb, M. J. et al. EMBO J. (1989) 8:2727-2736; Sherman, D. H. et al. EMBO J. (1989) 8:2717-2725; Fernandez-Moreno, M.A. et al. J.

Biol. Chem. (1992) 267:19278-19290). The "extender" units for the Type II PKSs are usually acetate (malonyl-CoA) units, and the presence of specific cyclases dictates the preferred pathway for cyclisation of the completed chain into an aromatic product (Hutchinson, C. R. and Fujii, I. Annu. Rev. Microbiol. (1995) 49:201-238). Hybrid polyketides have been obtained by the introduction of cloned Type II PKS gene-containing DNA into another strain containing a different Type II PKS gene cluster, for example by introduction of DNA derived from the gene cluster for actinorhodin, a blue-pigmented polyketide from *Streptomyces coelicolor*, into an anthraquinone polyketide-producing strain of *Streptomyces galileus* (Bartel, P. L. et al. J. Bacteriol. (1990) 172:4816-4826). Occasionally, unusual starter units are incorporated by Type II PKS, particularly in the biosynthesis of oxytetracycline, frenolicin and daunorubicin and in these cases a separate AT is used to transfer the starter unit to the PKS.

Fungal PKSs such as the 6-methylsalicylic acid or lovastatin PKS typically consist of a single multi-domain polypeptide which include most of the activities required for the synthesis of the polyketide portion of these molecules (Hutchinson C.R. and Fujii I. Annu. Rev. Microbiol. (1995) 49:201-238). Type II Fungal PKSs are

also known.

A number of mixed systems comprising polyketide synthase and nonribosomal peptide synthase modules have been identified including the epothilone and bleomycin biosynthetic clusters.

Although large numbers of therapeutically important polyketides have been identified, there remains a need to obtain novel polyketides that have enhanced properties or possess completely novel bioactivity. The complex polyketides produced by Type I PKSs are particularly valuable, in that they include compounds with known utility as anthelmintics, insecticides, anticancer, immunosuppressants, antifungal or antibacterial agents. Because of their structural complexity, such novel polyketides are not readily obtainable by total chemical synthesis, or by chemical modifications of known polyketides. Particular changes that are desired are changes to the carbon skeleton by altering the nature of the starter and/or extender unit(s) incorporated, changes to the oxidation level of the β -keto carbon and therefore the pattern of oxygen substituents by altering the series of reductive steps that occur after chain extension and changes to the post PKS "tailoring" steps which generally comprise hydroxylation, methylation or glycosylation of the polyketide molecule.

There is also a need to develop reliable and specific ways of deploying individual modules in practice so that all, or a large fraction, of hybrid PKS genes that are constructed, are viable and produce the desired polyketide product. Various strategies have been described to produce these hybrid PKSs particularly utilising recombinant DNA technology and denovo biosynthesis. There is a particular need to develop methods of manipulating these PKS in a manner that minimises the alteration to the PKS protein structure. Existing methods of achieving these manipulations sometimes produce hybrid PKS multienzymes which give the desired product at only 1% or less of the rate that the unmodified PKS produces product.

WO 93/13663 and WO 98/01571 describe novel methods of engineering PKSs. A well-established method of altering the nature of the extender unit used at any position in the polyketide molecule, particularly malonyl-, methylmalonyl- or ethylmalonyl-CoA is by domain substitution. For example, WO98/01546 and US patent 6,063,561 disclose methods of accomplishing this modification to form modified erythromycins. Novel polyketide molecules, in this case particularly novel erythromycins, are produced by the replacement of an entire AT domain-encoding DNA fragment on the

Saccharopolyspora erythraea chromosome with an equivalent heterologous AT domain-encoding fragment from another PKS cluster. It is well known to those skilled in the art that selection of the exact DNA/protein splice sites into
5 which to insert the heterologous domain requires detailed analysis of the corresponding DNA and protein sequences.

Different researchers choose to use splice sites at conserved, semi-conserved or non-conserved regions of the protein, or at sites either within or at the boundaries
10 of the AT domains. A further drawback of this technique is that it is hard to predict whether a particular heterologous domain will work in any given context. A domain that works successfully in one module may not work at all in an adjoining module or may produce polyketides
15 at a vastly reduced yield. Oliynyk, M. et al. (Chem. Biol. (1996) 3:833-839) and Ruan et al. (J. Bact. (1997) 179:6416-6425) have published studies that exchange a methylmalonyl-CoA specific AT domain for malonyl-CoA specific AT domains in modules of the erythromycin PKS.
20 Products were observed only for changes in modules 1 and 2, with module 2 at a vastly lowered yield. Stassi et al. (Proc. Natl. Acad. Sci. (1998) 95:7305-9) exchange the methylmalonyl-CoA specific AT of module 4 of the erythromycin PKS for an ethylmalonyl-CoA specific AT and

again product yield was low even after the addition of the crotonyl-CoA reductase gene thought to increase the supply of the required ethylmalonyl-CoA precursor. A possible reason for the limiting yields is the structural or mechanistic non-compatibility of a heterologous AT domain with the adjoining KS and ACP domains with which it must interact properly for efficient polyketide chain synthesis. Consequently, it is often necessary to try multiple domain swaps to achieve a novel polyketide-producing strain that displays adequate efficiency - a process made particularly arduous when these changes must be made by gene replacement on the chromosome through a two step double integration process. The introduction of splice sites at the DNA level is time consuming and technically challenging, requiring careful analysis to ensure the PKS protein coding reading frame is not disrupted. The introduction of restriction enzyme sites often requires changes at the amino acid level which lead to further PKS protein structure disruption and consequent loss of catalytic efficiency.

A method that could utilise the numerous techniques available for site directed mutagenesis to influence the AT substrate specificity with minimal disruption to the protein tertiary structure would be a valuable addition to the current techniques.

Changes to an active site have been shown to alter substrate specificity in other systems. For example, in an early study, Scrutton et al. (Nature (1990) 343:38-43) used site directed mutagenesis to switch the coenzyme
5 substrate specificity of a glutathione reductase. Identifying and changing a 'fingerprint' structural motif in the NADP+ binding domain they could convert the enzyme into one displaying a marked preference for NAD+. The techniques of directed evolution have been used to
10 improve/change enzyme catalytic function. Of many examples in the literature, Zhang et al. (PNAS (1997) 94:4504-4509) illustrate the conversion of a galactosidase to a fucosidase by these techniques. The resulting protein bears 6 mutations, of which 3 lie in,
15 or in close proximity to the active site.

Minor but directed changes to a PKS domain can make significant changes to its catalytic function. Patent application WO 00/00500 teaches that an extender ketosynthase domain is converted to a decarboxylating
20 (and hence loading) ketosynthase domain by site directed mutagenesis at the active site. US Patent numbers 6,004,787 and 6,066,721 and Jacobsen et al. Science (1997)277:367-369 describe the deletion of residues in the KS1 active site to inactivate this activity to allow
25 the production of novel polyketides by feeding of

synthetic precursors to the modified PKS.

Several studies have attempted to correlate the primary amino acid sequence of the AT to determine amino acids directly involved with the recognition of the appropriate substrate, and particularly the nature of the substrate side chain (i.e. the malonyl portion of the acyl-CoA thioester). Studies by Haydock et al. (FEBS Lett. (1995) 374:246-248) correlated the substrate specificity of malonyl- or methylmalonyl-CoA specific AT with a motif 11 amino acids upstream of the known active site. Comparisons between this motif and the protein structure of a known acyltransferase from *E. coli* fatty acid synthase allowed the authors to assess the proximity of the motif residues to the active site (and hence its ability to select the substrate). The authors acknowledged that *"this divergent region thus identified lies near the acyltransferase active site though not close enough to make direct contact with the substrate"*.

Other studies (Katz, L. Chem Rev. (1997) 97:2557-2575, Tang, L. et al., Gene (1998) 216:255-265) have correlated additional residues with a specific extender unit using these residues as a tool to predict the AT substrate specificity from a protein sequence derived from polyketide gene cluster sequencing projects. It has

remained unclear which residues have mechanistic importance. In only one case have regions within the PKS AT domain been exchanged in an attempt to swap AT specificity; patent application WO 00/01838 and Lau et al. Biochemistry (1999) 38:1643-51) implicated a 'hypervariable region' at the C-terminus of the AT domain in the selection of extender unit. These workers interchanged this 25-30 amino acid stretch and showed that this change was sufficient to alter the substrate specificity of the AT, concluding *"a short (23-35 amino acid) C-terminal segment present in all AT domains is the principal determinant of their substrate specificity. Interestingly its length and amino acid sequence vary considerably among the known AT domains. We therefore suggest that the choice of extender units by the PKS modules is influenced by a "hypervariable region", which could be manipulated via combinatorial mutagenesis to generate novel AT domains possessing relaxed or altered substrate specificity"*. Surprisingly, our structure molecular modelling studies indicate this region lies at a surface accessible region away from the active site and hence is unlikely to directly interact with (and hence directly select) the malonyl portion or the substrate

used. The effect on substrate specificity is therefore likely to be imprecise and due to more indirect effects via, for example, disruption of tertiary structure.

5 Disclosure of Invention

According to a first aspect of the present invention there is provided a method of synthesising a compound whereof at least a portion is the product of a polyketide synthase (PKS) enzyme complex or is derived from such a product, said PKS enzyme complex including at least one
10 acyltransferase (AT) domain. The method includes a step of providing said PKS enzyme complex in which said AT domain has been altered to change selectively a minor proportion of amino acid residues. The altered
15 residue(s) may comprise one or more motifs which are present in the active site pocket of the AT domain and which influence the substrate specificity of the AT domain, the alteration affecting the substrate specificity; and/or one or more residues of a motif which
20 influences the substrate specificity of the AT domain and which comprises a four-residue sequence corresponding to the YASH motif of the AT domain of the first module of DEBS, the alteration affecting the substrate specificity.

Synthesis is then effected by means of said PKS enzyme
25 complex to produce a compound or mixture of compounds

different from what could have been produced by means of a PKS enzyme in which said AT domain had not been altered.

The PKS enzyme complex may be at least part of a
5 modular type I PKS enzyme complex, or it may be derived from a type II PKS system, a fungal PKS system or a hybrid system comprising PKS and nonribosomal peptide synthase modules.

The present invention teaches that by altering a few
10 amino acid residues in the AT domain and particularly residues close to the AT active site comprising one or more residues of a short signature "motif" within the AT domain it is possible to influence the acylthioester selected by that AT domain. Novel polyketides can be
15 made by a modified PKS on which the signature motif on one or more modules is altered, e.g. being replaced with one associated with a different specificity for malonyl substrate. Furthermore, the invention provides a method of reducing the proportion of mixed polyketide products
20 that are occasionally found in natural systems due to non-specific incorporation of the incorrect extender units. Conversely, the invention provides a method of giving a mixed population of polyketide products thus increasing the diversity of polyketides produced by a
25 PKS.

The invention allows the preparation of a modified PKS by substitution of an existing amino acid residue motif in the AT that specifies incorporation of one of the common extender acylthioesters with another motif
5 found in another AT specifying an alternative acylthioester. This alters the substrate specificity of the polyketide synthase when it is expressed in a polyketide-producing organism.

The DNA sequences have been disclosed for numerous
10 Type I PKS gene clusters. Comprehensive sequence analysis of AT domains derived from Type I PKS modules responsible for the formation of macrolides, particularly erythromycin, rapamycin, avermectin, rifamycin, FK506, epothilone, tylosin, and niddamycin, ionophore
15 polyethers, particularly monensin, and polyenes, particularly nystatin, allowed us to identify amino acids that are characteristic of AT domains.

Figure 2 shows the sequence comparison of these AT domains. This sequence comparison has been generated in
20 a generally conventional way, employing a computer using a procedure that creates a multiple sequence alignment from a group of related sequences. We used a program called PileUp (Wisconsin Package, Genetics Computer Group (GCG), Madison, WI, USA), which creates a multiple
25 sequence alignment using simplification of the

progressive alignment method of Feng and Doolittle
(Journal of Molecular Evolution 25; 351-360 (1987)). The
method used is similar to the method described by Higgins
and Sharp (CABIOS 5; 151-153 (1989)). The program
5 executes a series of progressive, pairwise alignments
that allows a large number of sequences to be compared
together to form a final alignment throughout all the
sequences. Gaps can be inserted throughout individual
sequences to allow alignment of regions of strong
10 similarity. This is often required as strongly conserved
regions are often separated by more variable regions,
both in terms of numbers of amino acids and type of amino
acids. Different programs use different mathematical
algorithms to make these comparisons, resulting in
15 alignments that differ in minor ways. However, it can be
expected that regions of strong homology would still
align whatever alignment program is utilised. The
particular motifs that are discussed are marked.

These motifs include the conserved GQG motif that is
20 close to the start of the domain, the GHS motif that
contains the active site serine that covalently binds the
acyl chain prior to transfer to the ACP, and a LPTY motif
that is close to the end of the domain. Other residues
common to all ATs including an arginine, believed to
25 stabilise the carboxylate group of the acylthioester.

Further detailed sequence analysis allowed us to identify amino acid residues that differed between ATs responsible for the incorporation of malonyl-, methylmalonyl- and ethylmalonyl-CoA. Some of these amino acids or motifs
5 had been previously identified during the sequence analysis of the clusters as previously discussed. While these motifs could predict whether a malonyl- /methylmalonyl-CoA might be used they generally fail to show a difference between methylmalonyl- vs ethylmalonyl-
10 CoA or the other larger extender unit commonly used. We viewed this as an important requirement for identification of the most important and key residues involved in substrate recognition and consequently residues most suitable for alteration. Closer analysis
15 identified a string of four residues (location identified clearly in Figure 2) of which two residues are virtually invariant throughout all ATs, and two residues differ consistently depending on the extender unit.

Particularly, in the vast majority of ATs responsible for
20 recognition of malonyl-CoA the sequence of residues HAFH was identified. In the majority of ATs responsible for recognition of methylmalonyl-CoA the equivalent segment was substituted by residues YASH. In ATs responsible for ethylmalonyl-CoA or other similar sized CoA unit
25 incorporation the overall motif was different, less

conserved but generally displayed the sequence XAGH
(where X is most frequently but not limited to F, T, V or
H). We typically use the terms HAFH, YASH and TAGH to
describe these motifs with respect to malonyl-CoA,
5 methylmalonyl-CoA and ethylmalonyl/further CoA
specificity but use these terms herein to allow
substitutions in the motif, particularly at residue 1 as
described. Potential substitutions and the exact
location of the motif will be clear to those skilled in
10 the art by inspection of Figure 2 or similar sequence
analysis.

There are three possible methods to locate the
position of the motif within an AT sequence that does not
appear in Figure 2. It is likely a combination of the
15 methods will be used.

- I) By simple visual inspection and comparison of
the sequence to identify the motifs HAFH, YASH
or TAGH. Since substitutions of residue one
are often encountered a useful procedure is to
20 look for an alanine (A) separated by one amino
acid (usually F, S or G) from a histidine (H).
- II) By counting amino acids from the active site
serine. The start of the motif is typically
(but should not be limited to) between 90 and

100 amino acids downstream of the GHS active site motif.

III) By computer generated multiple alignment that allows the new sequence to be directly compared to the sequences and motifs we have annotated in Figure 2 or to other ATs.

It is preferable to use the third method as this allows the motif to be identified unequivocally when there are substitutions within the motif. This is particularly necessary in some of the more unusual types of AT in which one of the residues can be substituted by proline (P). The third method will also identify the motif when the number of residues between the motif and the AT active site serine differs significantly from the norm. The third method will also better identify the motif when the same or similar string of amino acids occurs elsewhere in the domain.

A particular feature of these motif residues is the relationship of the size of the third residue compared to the substrate selected. Hence, when malonyl-CoA is required the third residue is large (phenylalanine), when methylmalonyl-CoA is required this residue is intermediate (serine), and when ethylmalonyl-CoA is required this residue is small (glycine). The inverse relationship between substrate side chain size and this

third residue is particularly noteworthy. Interestingly, this relationship applies also when considering the incorporation of the more unusual extender units such as methoxymalonyl-CoA, required for some cycles of chain extension during production of for example FK506 (HAGH).

Currently, only a single example of an AT responsible for the incorporation of a five carbon-CoA unit has been disclosed. In this case the AT displays a different motif at this point, CPTH, in which only the histidine is conserved. The incorporation of a proline residue in the motif may be indicative of an AT specifying a larger substrate. Proline is also found in the motif in ATs that incorporate the larger unusual starter acids as seen in the case of avermectin and soraphen. Residues in and around this area, but lying in the active site of the AT domain define the specificity of the domain towards the substrate chosen.

Motifs that represent hybrids of motifs for malonyl- and methylmalonyl-CoA or methylmalonyl- and ethylmalonyl-CoA were identified. Particularly, epothilone module 3-expected HAFH or YASH (malonyl-CoA or methylmalonyl-CoA specific), found HASH or monensin module 5-expected TAGH (ethylmalonyl-CoA specific), found VAGH. Significantly, in both these cases the products of the PKS are a mixture due to the incorporation of 2 different extender units by

the module containing the hybrid motif, causing formation of monensins A and B and epothilones A and B. However, it is known that substrate supply is a significant determinant of the proportion of monensins A and B formed
5 (Liu, H. and Reynolds, K.A (1999) J. Bact. 181:6806-6813).

Many of the previously-proposed "predictive" motifs are unlikely to be the principal determinant of substrate specificity because they are not located in the active
10 site pocket. A particular requirement of any motif that can serve to distinguish between substrates is that it lies close to the active site and preferably within the substrate binding pocket. In this analysis we consider the substrate binding pocket to be the part of the pocket
15 that binds/recognises the malonyl portion of the acylthioester rather than necessarily the coenzyme A portion. In all probability some of the similarities previously identified by sequence analysis are due to evolutionary conservation rather than a mechanistic
20 requirement. In contrast the residues we have identified lie in or close to the substrate binding pocket. To assess the exact location of the motif in space we compared the protein sequence of ATs derived from Type I PKS with that of *E. coli* fatty acid malonyl-CoA:ACP
25 acyltransferase, for which there is a high resolution X-

ray crystal structure (Serre, L. et al., J. Biol. Chem. (1995) 270:12961-12964). While overall level of sequence similarity between these proteins is low, key residues (and particularly those with mechanistic importance) are conserved and the overall spatial arrangement of amino acids is expected to be conserved. Many groups have used this structure as a model AT and it is well known in the art that conservation of structure can be greater than the level of sequence conservation. Structural analysis showed that the identified motif would lie within the active site pocket opposite the active site serine and the arginine thought to be involved in binding the substrate carboxylate and close enough to the acyltransferase site to interact with the bound substrate side chain. The invariant histidine found in the motif is thought be part of a catalytic triad with the active site serine as is typically found in serine hydrolases (Serre et al, *Supra*). Figure 3 shows the position of the motif loop and important active site residues in the model AT structure.

Broadly the invention concerns modifying an AT domain by changing the four-residue sequence or motif responsible for selecting a substrate so that its specificity is altered. We may also change a small number of other residues close to the active site.

Generally the total number of residues changed is less than 5% of the residues of the AT.

The motif is the four-residue sequence corresponding to the YASH motif found at about residues 334-337 of the
5 AT domain of the first module of DEBS, numbering as shown in Fig. 2. It lies in the active site pocket. It typically starts 80-110, more particularly 90-100, amino acids downstream of the GHS active site motif.

In a preferred embodiment of this invention
10 polyketides of desired structure are produced by the replacement of an existing AT motif on a PKS with an alternative one responsible for selection of an alternative extender or starter unit, or responsible for an altered degree of selectivity (in most cases,
15 increased selectivity). This may be carried out in one or more of the modules encoding a PKS cluster. One type of embodiment is a PKS including two adjoining domains, which are "naturally" adjoining or otherwise coupled domains, wherein the first of them is an AT domain where
20 the four-residue motif has been altered to change its specificity, the AT domain acting to transfer a substrate to the second domain.

In one class of embodiments, this invention provides a PKS multienzyme or part thereof, or nucleic acid
25 (generally DNA) encoding it, said multienzyme or part

comprising a loading module and a plurality of extension modules for the generation of a polyketide, preferably selected from, macrolides, polyethers, or polyenes, wherein the loading or extension modules or at least one thereof contain a modified AT domain adapted to load and transfer an optionally substituted malonyl-CoA residue to (preferably) the ACP. The AT domain is preferably modified to alter its substrate specificity. This AT domain may differ from one naturally found in this position in the module only by the modification of a few amino acids lying in the active site. This modification comprises the exchange of all or part of a motif particularly but not limited to HAFH with YASH or TAGH or vice versa. Optionally, alterations to amino acids outside this sequence, but preferably lying close to the AT active site, are made.

A second class of embodiments provides a method of synthesising polyketides having a desired extension unit at any point around the polyketide molecule by providing a PKS multienzyme incorporating one or more modified AT domains and particularly but not limited to an AT domain possessing the motif HAFH or YASH or TAGH where these motifs replace the existing natural motif. Optionally, alterations to amino acids outside this sequence, but preferably lying close to the AT active site, are made.

A third class of embodiments provides a method of synthesising polyketides having a desired starter unit by providing a PKS multienzyme incorporating a modified AT domain in the loading module and particularly (but not
5 limited to) an AT domain possessing the motif HAFH or YASH or TAGH or a motif incorporating a proline residue where these motifs replace the existing natural motif. Optionally, alterations to amino acids outside this sequence, but preferably lying close to the AT active
10 site, are made. Preferentially, this AT will follow a KSQ domain but other loading systems are known in the art (e.g. AT-ACP). Patent application WO 00/00500 describes some of the known loading systems. The modification of the loading module can be combined with similar
15 modifications in other extension units.

A further class of embodiments provides a method of synthesising polyketides free of natural co-produced analogues and having a desired extender or loading unit by replacing an existing hybrid or alternative protein
20 motif with the sequences HAFH, YASH or TAGH. It is particularly useful to make this alteration in the epothilone or monensin PKS gene cluster.

In still further aspects this invention provides a method of synthesising a mixed population of polyketides
25 by providing a PKS multienzyme incorporating an AT with a

altered or hybrid motif, particularly, but not limited to HASH or VAGH. One particular utility of this method, though not limited to this utility, is the production of combinatorial libraries of compounds.

5 In a further aspect the PKS containing a modified AT domain may be spliced to a hybrid PKS produced for example as in WO 98/01546 and WO 98/01571 or WO 00/01827 or WO 00/00500. It is particularly useful to link such a modified PKS to gene assemblies that produce novel
10 derivatives of natural polyketides, for example 14-membered macrolides.

Each of these aspects and classes of embodiment may involve providing nucleic acid encoding the polyketide synthase multienzyme and introducing it into a organism
15 where it can be expressed. Suitable plasmids and host cells are described below. The polyketide synthase so produced or portions thereof may be isolated from the host cells by routine methods, though it is usually preferable not to do so. The host cells may also be
20 capable of producing the required acylthioester, eg. by producing ethylmalonyl CoA for example. It may be advantageous to remove the PKS from a strain with a particularly strong supply of an undesired acylthioester or express the altered PKS in a strain specifically
25 chosen to have a strong supply of a particular

acylthioester, or alternatively to develop media or growth conditions to enhance expression of the desired product. Conversely, such techniques could be used to promote formation of mixtures of products if so desired.

5 It may also be beneficial to supply chemical precursors to the desired acylthioesters in the media e.g. supply diethylethylmalonate or cyclobutane carboxylic acid etc.

The host cells may also be capable of modifying the initial PKS products, e.g. by carrying out all or some of
10 the biosynthetic modifications normal in the production of erythromycin (as shown in figure 4) and for other polyketides. Use may be made of mutant organisms such that some or all of the normal pathways are blocked, e.g. to produce products without one or more "natural" hydroxy
15 groups or methyl groups or sugar groups.

The invention should not be limited to the exact motifs described. We have described some of the known variations within the motif, particularly at residue 1 as can be determined by inspection of Figure 2 or by
20 inspection of similar sequence data. However other modifications can be envisaged; substitution of, for example, the phenylalanine in the malonyl-CoA motif by the similar sized tyrosine may still display the same selectivity. Similarly substitution of the small residue
25 glycine found in the motif responsible for ethylmalonyl-

CoA/other extender incorporation by for example but not limited to alanine. It is well known to those skilled in the art that these and other similar conservative substitutions frequently maintain the same selectivity.

5 Similarly the serine residue found in the motif for incorporation of methylmalonyl-CoA could be substituted by a residue intermediate in size and/or displaying a similar charge distribution.

The invention should not be limited to changes only
10 in this motif. Alterations to other residues around the AT domain may also be required to increase the level of specificity or catalytic efficiency, i.e. to increase the proportion or amounts of the desired products. These residues are preferentially close to the substrate
15 binding pocket. The requirement for these additional alterations will depend on the particular context or change desired. Particular residues to alter can be readily identified by inspection of Figure 2 or other similar sequence analysis data or alternatively by
20 analysis of the structural model.

Residues that may be altered in addition to the motif can be divided into two classes. Some of these residues may have been previously identified in the motifs used to predict the specificity of a motif (ie.
25 Haydock et al. (FEBS Lett. (1995) 374:246-248). These

residues are preferentially close to the substrate-binding pocket. These residues should not be limited to the particular examples described.

I) The first class of potential residues to change
5 includes residues close to the motif on the polypeptide chain. A particular example is the residue immediately after the 4 residue motif described in the present invention. In malonyl-CoA specific ATs this residue is generally serine (S), i.e. the protein sequence at this
10 point is generally HAFHS, whereas in methylmalonyl-CoA specific ATs this residue can be S but can also be T, G, or C for example. Thus to change a methylmalonyl-CoA specific AT to a malonyl-CoA specific AT by changing the signature motif it may be beneficial also to ensure that
15 the residue immediately after the motif is an S. Since this residue is close to the motif on the polypeptide chain it lies close to the substrate binding pocket.

II) The second class includes residues that are close to the motif or active site in space. These
20 residues are best identified by reference to the model AT structure described previously or another AT structure that may be subsequently derived. It is known to those skilled in the art that it is possible to thread related protein sequences into an existing structure by using
25 structure molecular modelling or related techniques.

Alternatively, an acylthioester may be modelled into the active site. These are the preferred methods, but often-simple inspection of the existing structure using the highly conserved motifs as a reference point gives a
5 reasonable approximation.

A particular example of a residue close in space to the motif that might be changed is the residue immediately after the GHS active site motif. In methylmalonyl-CoA specific ATs this residue is generally
10 glutamine (Q), i.e. the protein sequence at this point is GHSQ, whereas in malonyl-CoA specific ATs this residue is often V, I or L for example. Thus to change a malonyl-CoA specific AT to a methylmalonyl-CoA specific AT by changing the signature motif it may be beneficial also to
15 ensure that the residue immediately after the GHS motif is a Q. Since this residue is close to the active site serine it lies within the substrate-binding pocket.

A further example of a residue close in space that might be altered is the residue lying three residues
20 downstream of the GQG motif. In methylmalonyl-CoA specific ATs this residue is generally tryptophan (W), i.e. the protein sequence at this point is GQGXXW, whereas in malonyl-CoA specific ATs this residue is often R, H or T for example. Thus to change a malonyl-CoA
25 specific AT to a methylmalonyl-CoA specific AT by

changing the signature motif it may be beneficial also to ensure that this particular residue after the GQG motif is a W. Analysis of the model AT structure shows that the GQG motif lies close to the active site pocket and
5 consequently so does this tryptophan.

A further example of a residue close in space that might be altered is the residue 4 residues downstream from the conserved arginine referred to above, which is believed to stabilise the carboxylate group of the
10 acylthioester substrate. In malonyl-CoA specific ATs this residue downstream of the R is generally methionine (M), i.e. the protein sequence at this point is RXXMQ. In methylmalonyl-CoA specific ATs this residue is generally I or L, and in ethylmalonyl-CoA specific ATs it
15 is often W. Thus, for example, to change a methylmalonyl-CoA specific AT to a malonyl-CoA specific AT by changing the signature motif it may be beneficial also to ensure that this particular residue is a methionine. Analysis of the model AT structure shows
20 that this residue lies close to the active site pocket.

In further aspects the present invention provides vectors, such as plasmids or phages (preferably plasmids), including nucleic acids as defined in the above aspects and host cells particularly
25 *Saccharopolyspora* or *Streptomyces* species transformed

with such nucleic acids or constructs. It will be readily apparent to those skilled in the art that there are multiple molecular biological methods for achieving the desired alterations to the AT domain, particularly at the nucleic acid level, e.g. techniques of site directed mutagenesis or directed evolution. Suitable plasmid vectors and genetically engineered cells suitable for expression of PKS genes with modules incorporating an altered AT domain can readily be designed or selected by those skilled in the art. They include those described in WO 98/01546 as being suitable for expression of hybrid PKS genes of Type I. Examples of effective hosts are *Saccharopolyspora erythraea*, *Streptomyces coelicolor*, *Streptomyces avermitilis*, *Streptomyces griseofuscus*, *Streptomyces cinnamonensis*, *Streptomyces fradiae*, *Streptomyces longisporoflavus*, *Streptomyces hygroscopicus*, *Micromonospora griseorubida*, *Streptomyces lasaliensis*, *Streptomyces venezuelae*, *Streptomyces antibioticus*, *Streptomyces lividans*, *Streptomyces rimosus*, *Streptomyces albus*, *Amycolatopsis mediterranei*, and *Streptomyces tsukubaensis*. These include hosts in which SCP2*-derived plasmids are known to replicate autonomously, such as for example *S. coelicolor*, *S.*

avermitilis and *S. griseofuscus*; and other hosts such as *Saccharopolyspora erythraea* in which SCP2*-derived plasmids become integrated into the chromosome through homologous recombination between sequences on the plasmid insert and on the chromosome; and all such vectors which are integratively transformed by suicide plasmid vectors. A plasmid with an int sequence will integrate into a specific attachment site on the host's chromosome.

It is apparent to those skilled in the art that the overall sequence similarity between nucleic acids encoding comparable AT domains from Type I PKSs is sufficiently high and the domain organisation of different Type I PKSs so consistent between different polyketide-producing organisms, that the processes for obtaining novel hybrid polyketides described will be generally applicable to all natural modular Type I PKSs or their derivatives.

The present invention will now be illustrated, but is not intended to be limited, by means of some examples.

Amino acids have been defined throughout by their standard one letter codes as follows. A-alanine, R-arginine, N-asparagine, D-aspartic acid, C-cysteine, Q-glutamine, E-glutamic acid, G-glycine, H-histidine, I-isoleucine, L-leucine, K-lysine, M-methionine, F-

phenylalanine, P-proline, S-serine, T-threonine, W-tryptophan, Y-tyrosine and V-valine.

Brief Description of Drawings

5 Figure 1 is a diagram showing the functioning of 6-deoxyerythronolide B synthase (DEBS), a modular PKS producing 6-deoxyerythronolide B, a precursor of erythromycin A.

10 Figure 2 gives the amino acid sequence comparison of the AT domains of representative Type I PKS gene clusters. The motifs GQG, GHS and LPTY are marked at the base of the figure along with the arginine and the motif defined in the invention as defining specificity. The abbreviations used at the side to define the PKS used
15 are: ave: avermectin, debs: erythromycin, epo: epothilone, sor: soraphen, fkb: FK506, rap: rapamycin, tyl: tylosin, mon: monensin, nid: niddamycin, nys: nystatin, rif: rifamycin. The numbers represent the module number. The letter a at the end of the
20 designation indicates malonyl-CoA specific AT, the letter p indicates methylmalonyl-CoA specific AT, and the letter b indicates ethylmalonyl-CoA specific AT. Further types of AT with unusual or ill-defined AT specificity are indicated with letter x. Due to the numbers of sequences
25 considered, in the pileup each section of 50 amino acids

spreads over two pages. The sequences of the monensin ATs are unpublished. They are set out in PCT/GB00/02072.

Figure 3 shows a three-dimensional representation of the active site of the *E. coli* acyltransferase. The spatial arrangement of the motifs described in the text are shown by arrows and the atoms shown in bold.

Figure 4 shows the enzymatic steps that convert 6-deoxyerythronolide B into erythromycin A in *Saccharopolyspora erythraea*.

Figure 5 shows the DNA sequence from the monensin PKS encoding the loading AT used in Example 8.

Modes for Carrying Out the Invention

Example 1

Construction of plasmid pHP41

Plasmid pHP41 is a pCJR24-based plasmid containing the DEBS1 PKS gene comprising a loading module, the first and second extension modules of DEBS and the chain terminating thioesterase. The motif YASH of the AT domain of first module has been altered to HAFH. Plasmid pHP41 was constructed by several intermediate plasmids as follows. Plasmid pD1AT2 (Oliynyk, M. et al. Chem. Biol. (1996) 3:833-839) was digested with NdeI and XbaI. A ~11kbp fragment was isolated by gel electrophoresis and

the DNA purified from the gel. This fragment was ligated into pCJR24 (Rowe, C.J. et al. Gene (1998) 216:215-223) that had been linearised by digestion with *NdeI* and *XbaI* and treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E. coli* DH10B cells and individual clones checked for the desired plasmid pCJR26. Plasmid pCJR26 was identified by restriction pattern. pCJR26 was transformed into *E. coli* strain ET12567 (McNeil, D.J. et al. Gene (1992) 111:61-68) and an individual colony grown overnight to isolate demethylated DNA. This DNA was linearised using *MscI* and *AvrII* and the ~13kb fragment (Fragment A) isolated by gel electrophoresis and purification from the gel.

A DNA segment of the *eryAI* gene (start nucleotide 45368, end nucleotide 34734) from *S. erythraea* extending from nucleotide 42104 to nucleotide 41542 was amplified by PCR using the following oligonucleotide primers; 5'-TTTTTTTGGCCAGGGTTGGCAGTGGGCGGGCA-3' and 5'-TTTTTACGGCCAGCCGCTTGGCGCGGAT-3'. The DNA from a plasmid designated pCJR65 derived from pCJR24 and DEBS1TE was used as a template. The design of the primers introduced a *MscI* site at nucleotide 42105 and the second primed across a *BstXI* site at position 41546. The 574bp PCR

product was treated with T4 polynucleotide kinase and ligated to plasmid pUC18 that had been linearised by digestion with *Sma*I and then treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E. coli* DH10B and individual clones checked for the presence of the desired plasmid pHP39. Plasmid pHP39 was identified by restriction pattern and sequence analysis. Demethylated DNA was produced by transforming *E. coli* strain ET12567 with plasmid DNA.

10 The resulting DNA was linearised by digestion with *Msc*I and *Bst*XI and the resulting 552bp fragment (Fragment B) isolated by gel electrophoresis and purified from the gel. A DNA segment of the *eryA*I gene from *S. erythraea* extending from nucleotide 41557 to nucleotide 41120 was amplified by PCR using the following oligonucleotide primers; 5'-CGGTGCCTAGGTGCACCGACTCCCAGTCC-3 5'-

15 TTTTTC AAGCGGCTGGCCGTGGACCACGCGTTCCACTCCTCGCACGTCGAGACGAT -3'. DNA from plasmid pCJR65 was used as a template. The design of the primers introduced an *Avr*II site at

20 nucleotide 41125 and the second primed across a *Bst*XI site at nucleotide 41557 and mutated the amino acid sequence YASH to HAFH (encoded by nucleotides 41537-41526). The 442bp PCR product was treated with T4 polynucleotide kinase and ligated to plasmid pUC18 that

had been linearised by digestion with *Sma*I and then treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E. coli* DH10B and individual clones checked for the presence of the
5 desired plasmid pHP40. Plasmid pHP40 was identified by restriction pattern and sequence analysis. Plasmid pHP40 was linearised by digestion with restriction enzymes *Avr*II and *Bst*XI, and a 427bp fragment (Fragment C) isolated by gel electrophoresis and purified from the
10 gel. Fragments A, B, and C were ligated together and the resulting ligation mixture used to transform electrocompetent *E. coli* DH10B. Individual clones were checked for the presence of an insert derived from DEBS1. The resulting plasmid was designated pHP41. Sequence
15 analysis was used to confirm the clone contained the correct motif HAFH.

Example 2

Construction of *S. erythraea* NRRL2338 JC2/pHP41 and 20 production of triketides

S. erythraea NRRL2338 JC2 contains a deletion of the *eryAI*, *eryAII* and *eryAIII* apart from the TE (Rowe, C.J. et al. Gene 216, 215-223). Plasmid pHP41 was used to transform *S. erythraea* NRRL2338 JC2 protoplasts using the

TE as a homology region. Thiostrepton resistant colonies were selected on R2T20 agar containing 40 µg/ml thiostrepton. *S. erythraea* NRRL2338 JC2 (pHP41) was plated onto SM3 agar (see patent application WO 00/01827) containing 40 µg/ml thiostrepton and allowed to grow for 11 days at 30°C. Approximately 1cm² of the agar was homogenised and extracted with a mixture of 1.2ml ethyl acetate and 20 µl formic acid. The solvent was decanted and removed by evaporation and the residue dissolved in methanol and analysed by GC/MS. The major products were identified by comparison with authentic standards (Oliynyk, M. et al. Chem. Biol. (1996) 3:833-839) as triketide lactones (2S,3R,5R)-2-methyl-3,5-dihydroxy-n-hexanoic δ-lactone (AAP, i.e. Acetate, Acetate, Propionate incorporation), (2S,3R,5R)-2-methyl-3,5-dihydroxy-n-heptanoic δ-lactone (PAP), (2R,3S,4S,5R) 2, 4-dimethyl-3,5-dihydroxy-n-heptanoic δ-lactone (PPP) and (2R,3S,4S,5R) 2, 4-dimethyl-3,5-dihydroxy-n-hexanoic δ-lactone (APP). These products were identified as their ammonium adducts corresponding to exact mass 144, 158, 172 and 158. Four products were produced because in this strain, and under the conditions of the experiment the loading module loads both acetate and propionate and the modified AT loads malonyl-CoA and methylmalonyl-CoA.

Only three triketide lactone peaks could be observed in the GC/MS spectra under standard conditions, this was due to the co-elution of the equivalent mass APP and PAP compounds. An isocratic gradient was used to verify this peak was comprised of two components. In further sets of experiments *S. erythraea* JC2 (pHP41) was used to inoculate 5ml TSB containing 5 µg/ml thiostrepton. After three days growth 1.5ml of this culture was used to inoculate 25ml SM3 media containing 5 µg/ml thiostrepton in a 250ml flask. The flask was incubated at 30 °C, 250rpm for 6 days. At this time the supernatant was adjusted to pH3.0 with formic acid and extracted twice with an equal volume of ethyl acetate. The solvent was removed by evaporation and the residue analysed by GC/MS. In each experiment we could identify the 4 products AAP, PAP, PPP and APP but the absolute ratios and quantities were variable, presumably depending on exact media and growth conditions within each flask (figure 6).

20 Example 3

Construction of *S. erythraea* NRRL2338 (pHP41) and its use to produce 12-desmethyl erythromycin B.

Plasmid pHP41 was used to transform *S. erythraea* NRRL2338 protoplasts. Thiostrepton resistant colonies

were selected on R2T20 agar containing 40 µg/ml
thiostrepton. Several clones were tested for the
presence of pHP41 integrated into the chromosome by
Southern blot hybridisation of their genomic DNA with DIG
5 labelled vector DNA. A clone with a correctly integrated
copy of pHP41 was identified in this way. *S. erythraea*
NRRL2338 (pHP41) was used to inoculate 5ml TSB containing
5 µg/ml thiostrepton. After three days growth 1.5ml of
this culture was used to inoculate 25ml EryP media (see
10 patent application WO 00/00500) containing 5 µg/ml
thiostrepton in a 250ml flask. The flask was incubated
at 30 °C, 250rpm for 6 days. At this time the supernatant
was adjusted to pH9.0 with ammonia and extracted twice
with an equal volume of ethyl acetate. The solvent was
15 removed by evaporation and the residue analysed by
HPLC/MS. A peak of molecular mass m/z (M+H)=704 was
observed required for C-12 desmethyl erythromycin B in
addition to a peak corresponding to erythromycin A
(M+H)=734. Other peaks corresponding to partially
20 processed erythromycin intermediates could be identified.

Example 4

Construction of plasmid pHP048

Plasmid pHP048 is a pCJR24-based plasmid containing the
25 DEBS1 PKS gene comprising a loading module, the first and

second extension modules of DEBS1 and the chain terminating thioesterase. The motif YASH of the AT domain of first module has been altered to HASH. Plasmid PHP048 was constructed by several intermediate plasmids as follows.

A DNA segment of the *eryAI* gene from *S. erythraea* extending from nucleotide 41557 to nucleotide 41120 was amplified by PCR using the following oligonucleotide primers; 5'-CGGTGCCTAGGTGCACCGACTCCCAGTCC-3' and 5'-
10 TTTTTCGAAGCGGCTGGCCGTGGACCACGCGTCGCACTCCTCGCACGTCGAGACGAT-3'. The DNA from plasmid pCJR65 was used as a template. The design of the primers introduced a *AvrII* site at nucleotide 41125 and the second extended to a *BstXI* site at nucleotide 41557, also mutated the amino acid sequence
15 YASH (encoded by nucleotides 41537-41526) to HASH. The 442bp PCR product was treated with T4 polynucleotide kinase and ligated to plasmid pUC18 that had been linearised by digestion with *SmaI* and then treated with alkaline phosphatase. The ligation mixture was used to
20 transform electrocompetent *E. coli* DH10B and individual clones checked for the presence of the desired plasmid PHP022. Plasmid PHP022 was identified by restriction pattern and sequence analysis. Plasmid PHP022 was linearised by digestion with restriction enzymes *AvrII*

and *Bst*XI, and the fragment (Fragment D) isolated by gel electrophoresis and purified from the gel. Fragment D was ligated with Fragments A and B described previously and the resulting ligation mixture used to transform
5 electrocompetent *E. coli* DH10B. Individual clones were checked for the presence of an insert derived from DEBS1. The resulting plasmid was designated pHP048. Sequence analysis was used to confirm the clone contained the correct motif HASH.

10

Example 5

Construction of *S. erythraea* NRRL2338 JC2 (pHP048)

and its use to produce triketides

S. erythraea NRRL2338 JC2 contains a deletion of the
15 *eryAI*, *eryAII* and *eryAIII* apart from the TE (Rowe, C.J. *et al.* Gene 216, 215-223). Plasmid pHP048 was used to transform *S. erythraea* NRRL2338 JC2 protoplasts using the TE as a homology region. Thiostrepton resistant colonies were selected on R2T20 agar containing 40µg/ml
20 thiostrepton. *S. erythraea* JC2 (pHP048) was used to inoculate 5ml TSB containing 5 µg/ml thiostrepton. After three days growth 1.5ml of this culture was used to inoculate 25ml SM3 media containing 5 µg/ml thiostrepton in a 250ml flask. The flask was incubated at 30 °C,

250rpm for 6 days. At this time the supernatant was adjusted to pH3.0 with formic acid and extracted twice with an equal volume of ethyl acetate. The solvent was removed by evaporation and the residue analysed by GC/MS.

5 A mixture of products were identified as their ammonium adducts corresponding to the AAP, PAP, APP and PPP triketide lactones as described in example 2. In this example, under the media/growth conditions described the PKS with the HASH change is more catalytically active

10 than the HAFH change (example 2) as judged by total amounts of triketide lactone produced, however in this case the modified PKS appears to display lower selectivity towards acetate as judged by the ratio of AAP to PPP triketide lactone.

15

Example 6

Construction of plasmid pHP47

Plasmid pHP47 is a pCJR24-based plasmid containing

20 the DEBS1 PKS gene comprising a loading module, the first and second extension modules of DEBS1 and the chain terminating thioesterase. The motif YASH of the AT domain of first module has been altered to VAGH. Plasmid pHP47 was constructed by several intermediate plasmids as

25 follows.

A DNA segment of the *eryAI* gene from *S. erythraea* extending from nucleotide 41557 to nucleotide 41120 was amplified by PCR using the following oligonucleotide primers; 5'-CGGTGCCTAGGTGCACCGACTCCCAGTCC-3' and 5'-

5 TTTTTC AAGCGGCTGGCCGTGGACGTCGCGGGGCACTCCTCGCACGTCGAGACGAT
-3'. The DNA from plasmid pCJR65 was used as a template.

The design of the primers introduced a *AvrII* site at nucleotide 41125 and the second extended to a *BstXI* site at nucleotide 41557, also mutated the amino acid sequence

10 YASH (encoded by nucleotides 41537-41526) to VAGH. The 442bp PCR product was treated with T4 polynucleotide kinase and ligated to plasmid pUC18 that had been linearised by digestion with *SmaI* and then treated with alkaline phosphatase. The ligation mixture was used to

15 transform electrocompetent *E. coli* DH10B and individual clones checked for the presence of the desired plasmid pHP46. Plasmid pHP46 was identified by restriction pattern and sequence analysis. Plasmid pHP46 was linearised by digestion with restriction enzymes *AvrII*

20 and *BstXI*, and the fragment (Fragment E) isolated by gel electrophoresis and purified from the gel. Fragment E was ligated with Fragments A and B described previously and the resulting ligation mixture used to transform electrocompetent *E. coli* DH10B. Individual clones were

checked for the presence of an insert derived from DEBS1.

The resulting plasmid was designated pHP47. Sequence analysis was used to confirm the clone contained the correct motif VAGH.

5

Example 7

Construction of plasmid pLS007

Plasmid pLS007 contains the crotonyl-CoA reductase (CCR) gene from *S. cinnamonensis* that is believed to

10 influence the level of ethylmalonyl-CoA within the cell.

Plasmid pSG142 (Gaisser et al. Mol. Microbiol. (2000) 36

391-401) places genes under the control of the actI

promoter and can be used to integrate either in the right

hand side of the erythromycin gene cluster or in the act

15 promoter region of a previously transformed actinomycete.

Two oligonucleotide primers; 5'-

GGCAAACATATGAAGGAAATCCTGGACGCG-3' and 5'-

TCCGCGGATCCTCAGTGCGTTCAGATCAGTGC-3' were used to amplify

the *S. cinnamonensis* CCR gene using genomic DNA as

20 template. The design of the primers incorporated NdeI

and BamHI restriction sites to facilitate cloning. The

1.4kb PCR product was isolated by gel electrophoresis and

purified from the gel and ligated with pSG142 that had

been digested with NdeI and BglII. The resulting

ligation mixture was used to transform electrocompetent *E. coli* DH10B cells. Plasmid pLS003 was identified by restriction analysis and sequencing to ensure errors were not introduced during amplification. A discrepancy with the published sequence was identified. However, further analysis by comparison with other published CCR protein sequences indicated pLS003 was correct. Plasmid pLS003 was digested with *NdeI* and *XbaI* and the resulting 4.5kb fragment (fragment F) isolated by gel electrophoresis and purified from the gel. This fragment was ligated to pLSB2 a derivative of pKC1132 containing the actI/actII promoter region behind an *NdeI* site. Plasmid pLSB2 was digested with *NdeI* and *XbaI* and the resulting ~4kb fragment (Fragment G) purified by gel electrophoresis and purified from the gel. Fragments F and G were ligated together and the resulting ligation mixture was used to transform electrocompetent *E. coli* DH10B cells. Plasmid pLS007 was identified by restriction analysis.

20 Example 8

Construction of *S. erythraea* NRRL2338 JC2 (pHP47/pLS007) and its use to produce triketides

S. erythraea NRRL2338 JC2 contains a deletion of the *eryAI*, *eryAII* and *eryAIII* apart from the TE (Rowe, C.J.

et al. Gene 216, 215-223). Plasmid pHP47 was used to transform *S. erythraea* NRRL2338 JC2 protoplasts using the TE as a homology region. Thiostrepton resistant colonies were selected on R2T20 agar containing 40 µg/ml

5 thiostrepton. PLS007 was used to transform protoplasts of *S. erythraea* NRRL2338 JC2 (pHP47), thiostrepton and apramycin resistant clones were selected on R2T20 agar containing 40 µg/ml thiostrepton and 50 µg/ml apramycin plus 10mM magnesium chloride and the resistance markers

10 verified by plating on tapwater media containing the same antibiotics. *S. erythraea* NRRL2338 JC2 (pHP47/pLS007) was used to inoculate 5ml TSB containing 5 µg/ml thiostrepton and 50 µg/ml apramycin. After three days growth 1.5ml of this culture was used to inoculate 25ml SM3 media

15 containing 5 µg/ml thiostrepton and 50 µg/ml apramycin in a 250ml flask. The flask was incubated at 30°C, 250rpm for 6 days. At this time the supernatant was adjusted to pH3.0 with formic acid and extracted twice with an equal volume of ethyl acetate. The solvent was removed by

20 evaporation and the residue analysed by GC/MS. In this experiment amounts of triketide product were lower but a mixture of products could be identified as their ammonium adducts corresponding to exact masses 158 172 and 186.

Example 9Construction of *S. erythraea* NRRL2338 (pHP47) and its use to produce erythromycins.

Plasmid pHP47 was used to transform *S. erythraea*
5 NRRL2338 protoplasts. Thiostrepton resistant colonies
were selected on R2T20 agar containing 40 µg/ml
thiostrepton. *S. erythraea* NRRL2338 (pHP47) was used to
inoculate 5ml TSB containing 5 µg/ml thiostrepton. After
three days growth 1.5ml of this culture was used to
10 inoculate 25ml EryP media containing 5 µg/ml thiostrepton
in a 250ml flask. The flask was incubated at 30°C, 250rpm
for 6 days. At this time the supernatant was adjusted to
pH9.0 with ammonia and extracted twice with an equal
volume of ethyl acetate. The solvent was removed by
15 evaporation and the residue analysed by HPLC/MS. Peaks
of mass m/z ($M+H$) = 734 corresponding to erythromycin A
were observed.

Example 1020 Construction of plasmid pSGK051

Plasmid pSGK051 is a pPFL43 based plasmid (WO
00/00500). The motif HAFH of the AT domain of the
loading domain has been altered to YASH. Plasmid pSGK051
was constructed by several intermediate plasmids as

follows.

Plasmid pPFL43 was linearised by digestion with restriction enzymes *Nco*I and *Not*I and a 858bp fragment (Fragment Q) isolated by gel electrophoresis and purified
5 from the gel.

A DNA segment of the monensin loading domain from nucleotide 16360-17366 (see figure 5 and PCT/GB00/02072) was amplified by PCR using the following oligonucleotide primers; 5'-
10 GGGGACGCGGCCGCAAGGCCACACCTGAAGGTCAGCTACGCCTCCCACTCCCCGC
ACATGGACCCCAT-3' and 5'-GGCTAGCGGGTCCTCGTCCGTGCCGAGGTCA-
3'. The design of the primers amplified across a *Not*I site at nucleotide 16367 and changed the amino acid sequence HAFH to YASH at nucleotides 16398-16409, the
15 second introduced a *Nhe*I site equivalent to that in pPFL43. The DNA from plasmid pPFL43 was used as a template. The 1006bp PCR product was treated with T4 polynucleotide kinase and ligated to plasmid pUC18 that had been linearised by digestion with *Sma*I and treated
20 with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E. coli* DH10B and individual clones checked for the presence of the desired plasmid pCSAT9. Plasmid pCSAT9 was identified by restriction pattern and sequence analysis. Plasmid

pCSAT9 was linearised by digestion with restriction enzymes NotI and NheI and a 995bp fragment (Fragment R) isolated by gel electrophoresis and purified from the gel. Plasmid pPFL43 was digested with NcoI and NheI to
5 remove a 1.8kb fragment and the larger fragment (Fragment S) isolated by gel electrophoresis and purified from the gel. Fragments Q, R and S were ligated together and the resulting ligation mixture used to transform electrocompetent *E. coli* DH10B. Individual clones were
10 checked for the desired plasmid pSGK051. The resulting plasmid was analysed by restriction digest and sequenced to confirm the presence of the correct motif YASH.

Example 11

15 Construction of *S. erythraea* NRRL2338 JC2/pSGK051 and production of triketides.

Plasmid pSGK051 was used to transform *S. erythraea* NRRL2338 JC2 protoplasts using the TE as a homology region. Thiostrepton resistant colonies were selected on
20 R2T20 agar containing 40 µg/ml thiostrepton. *S. erythraea* NRRL2338 JC2 (pSGK051) was plated onto R2T20 agar containing 40 µg/ml thiostrepton and allowed to grow for 11 days at 30°C. Approximately 1cm² of the agar was homogenised and extracted with a mixture of 1.2ml ethyl

acetate and 20 μ l formic acid. The solvent was decanted and removed by evaporation and the residue dissolved in methanol and analysed by GC/MS. The major products were identified by comparison with authentic standards as
5 triketide lactones (2S,3R,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-heptanoic δ -lactone and (2S,3R,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-hexanoic δ -lactone.

Example 12

10 Construction of *S. erythraea* NRRL2338 (pSGK051) and its use to produce erythromycins.

Plasmid pSGK051 was used to transform *S. erythraea* NRRL2338 protoplasts. Thiostrepton resistant colonies were selected on R2T20 agar containing 40 μ g/ml
15 thiostrepton. *S. erythraea* NRRL2338 (pSGK051) was plated onto R2T20 agar containing 40 μ g/ml thiostrepton and allowed to grow for 10 days at 30°C. Approximately 2cm² of the agar was homogenised and extracted with a mixture of 1.2ml ethyl acetate and 20 μ l dilute ammonia. The
20 solvent decanted and was removed by evaporation and the residue analysed by HPLC/MS. Peaks of mass m/z (M+H)=734 and 720 could be observed alongside likely products of incomplete processing. Comparison to authentic standards proved the compounds produced were erythromycin A and 13-

methyl erythromycin A.

CLAIMS:

1. A method of synthesising a compound whereof at least a portion is the product of a polyketide synthase (PKS) enzyme complex or is derived from such a product, said PKS enzyme complex including at least one acyltransferase (AT) domain; said method comprising the steps of (i) providing said PKS enzyme complex in which said AT domain has been altered to change selectively a minor proportion of amino acid residues, the altered residue(s) comprising one or more residues of one or more motifs which are present in the active site pocket of the AT domain and which influence the substrate specificity of the AT domain, the alteration affecting the substrate specificity; and (ii) effecting synthesis by means of said PKS enzyme complex to produce a compound or mixture of compounds different from what could have been produced by means of a PKS enzyme in which said AT domain had not been altered.
2. A method according to claim 1 wherein said motif comprises a four-residue sequence corresponding to the YASH motif of the AT domain of the first module of DEBS.
3. A method of synthesising a compound whereof at least a portion is the product of a polyketide synthase

(PKS) enzyme complex or is derived from such a product, said PKS enzyme complex including at least one acyltransferase (AT) domain; said method comprising the steps of (i) providing said PKS enzyme complex in which
5 said AT domain has been altered to change selectively a minor proportion of amino acid residues, the altered residue(s) comprising one or more residues of a motif which influences the substrate specificity of the AT domain and which comprises a four-residue sequence
10 corresponding to the YASH motif of the AT domain of the first module of DEBS, the alteration affecting the substrate specificity; and (ii) effecting synthesis by means of said PKS enzyme complex to produce a compound or mixture of compounds different from what could have been
15 produced by means of a PKS enzyme in which said AT domain had not been altered.

4. A method according to claims 1, 2 or 3 wherein said motif was located by a) determining the sequence of the AT domain and b) performing sequence alignment with a
20 plurality of sequences of other AT domains.

5. A method according to any preceding claim wherein the PKS enzyme complex is at least part of a modular type I PKS enzyme complex.

6. A method according to any preceding claim wherein said alteration of the AT domain affects less than 5% of the residues.

7. A method according to any preceding claim
5 wherein said alteration alters a motif selected from XAFH, XASH, and XAGH and/or creates such a motif.

8. A method according to claim 7 wherein the motif is XAGH and X is selected from F, T, V and H.

9. A method according to claim 7 wherein the motif
10 is XAFH and X is H.

10. A method according to claim 7 wherein the motif is XASH and X is selected from Y, H, W and V.

11. A method according to any of claims 1-10 wherein said alteration produces or alters a motif
15 containing proline.

12. A method according to any preceding claim wherein in addition to the alteration to one or more residues of said motif(s), one or more additional residues in or adjacent the substrate binding pocket have
20 been altered.

13. A method according to claim 12 wherein said additional altered residue(s) comprise one or more of a) the residue immediately downstream of the motif, b) the residue three residues downstream from the GQG motif, c)
25 the residue immediately downstream of the GHS motif, and

d) the residue four residues downstream of the conserved arginine residue.

14. A method according to any preceding claim wherein the alteration produces a motif specific for
5 malonyl-CoA and the motif is followed by S which was produced by alteration if not already present.

15. A method according to any of claims 1-13 wherein the alteration produces a motif specific for methylmalonyl-CoA and the motif is followed by S, G, C or
10 T which was produced by alteration if not already present.

16. A method according to any of claims 1-13 wherein the alteration produces a motif specific for methylmalonyl-CoA, and the residue following the GHS
15 motif in the active site is Q which was produced by alteration if not already present.

17. A method according to any of claims 1-13 wherein the alteration produces a motif specific for malonyl-CoA, and the residue following the GHS motif in
20 the active site is V, I or L which was produced by alteration if not already present.

18. A method according to any of claims 1-13 wherein the alteration produces a motif specific for methylmalonyl-CoA, and the residue 3 residues downstream

of the GQG motif is W which was produced by alteration if not already present.

19. A method according to any of claims 1-13 wherein the alteration produces a motif specific for
5 malonyl-CoA, and the residue 3 residues downstream of the GQG motif is R, H or T which was produced by alteration if not already present.

20. A method according to any of claims 1-13 wherein the alteration produces a motif specific for
10 malonyl-CoA and the residue 4 residues downstream of the conserved R as found as residue 252 in the first module of DEBS is M which was produced by alteration if not already present.

21. A method according to any of claims 1-13
15 wherein the alteration produces a motif specific for methylmalonyl-CoA and the residue 4 residues downstream of the conserved R as found as residue 252 in the first module of DEBS is I or L which was produced by alteration if not already present.

20 22. A method according to any of claims 1-13 wherein the alteration produces a motif specific for ethylmalonyl-CoA and the residue 4 residues downstream of the conserved R as found as residue 252 in the first module of DEBS is W which was produced by alteration if
25 not already present.

23. A method according to any preceding claim wherein the AT domain has an active site with a GHS motif, and said motif which is altered starts 80-110 residues downstream of said GHS motif.

5 24. A method according to any preceding claim wherein said step (i) of providing said PKS enzyme complex comprises providing a nucleic acid sequence encoding said complex and effecting expression thereof.

25. A method according to claim 24 wherein
10 expression is effected in an organism capable of producing polyketides.

26. A method according to claim 24 or claim 25 wherein said nucleic acid sequence has been subjected to site directed mutagenesis so that it encodes said altered
15 AT domain.

27. A method according to claim 24, 25 or 26 wherein the AT domain prior to alteration is naturally expressed in a first organism and the altered AT is expressed in a second organism which is better able than
20 the first organism to supply a substrate for which the alteration has increased specificity and/or which is less well able than the first organism to supply a substrate for which the alteration has reduced specificity.

28. A method according to any preceding claim
25 wherein said PKS includes said AT domain and a second

domain which is naturally coupled thereto prior to the alteration thereof to receive a substrate transferred to it by the AT; and the alteration causes the AT to act to transfer a different substrate to the second domain.

5 29. A method according to any preceding claim wherein said PKS includes said AT domain and its natural cognate ACP domain which, prior to the alteration, is adapted to receive a substrate transferred to it by the AT; and the alteration causes the AT to act to transfer a
10 different substrate to said cognate ACP domain.

30. A method according to any preceding claim wherein said PKS including the altered AT domain is spliced to a hybrid PKS.

31. A polyketide compound or derivative thereof or
15 compound whereof a portion is a polyketide or derivative thereof, which compound is obtainable by a method according to any preceding claim wherein the compound differs from a compound resulting from synthesis effected by means of said PKS enzyme complex without the
20 alteration of said AT domain.

32. Nucleic acid encoding a PKS enzyme complex including an altered AT domain as defined in any of claims 1-30.

33. A vector including a nucleic acid according to
25 claim 32.

34. A host organism containing nucleic acid according to claim 32 and able to express the PKS enzyme complex.

35. A host organism according to claim 34 which is adapted to synthesise a compound whereof at least a portion is a polyketide resulting from the action of the PKS enzyme complex.

36. A method of synthesising a polyketide synthase (PKS) enzyme complex, said PKS enzyme complex including at least one acyltransferase (AT) domain; said method comprising altering said AT domain to change selectively a minor proportion of amino acid residues, the altered residue(s) comprising one or more residues of one or more motifs which are present in the active site pocket of the AT domain and which influence the substrate specificity of the AT domain, the alteration affecting the substrate specificity.

37. A method according to claim 36 wherein said motif comprises a four-residue sequence corresponding to the YASH motif of the AT domain of the first module of DEBS.

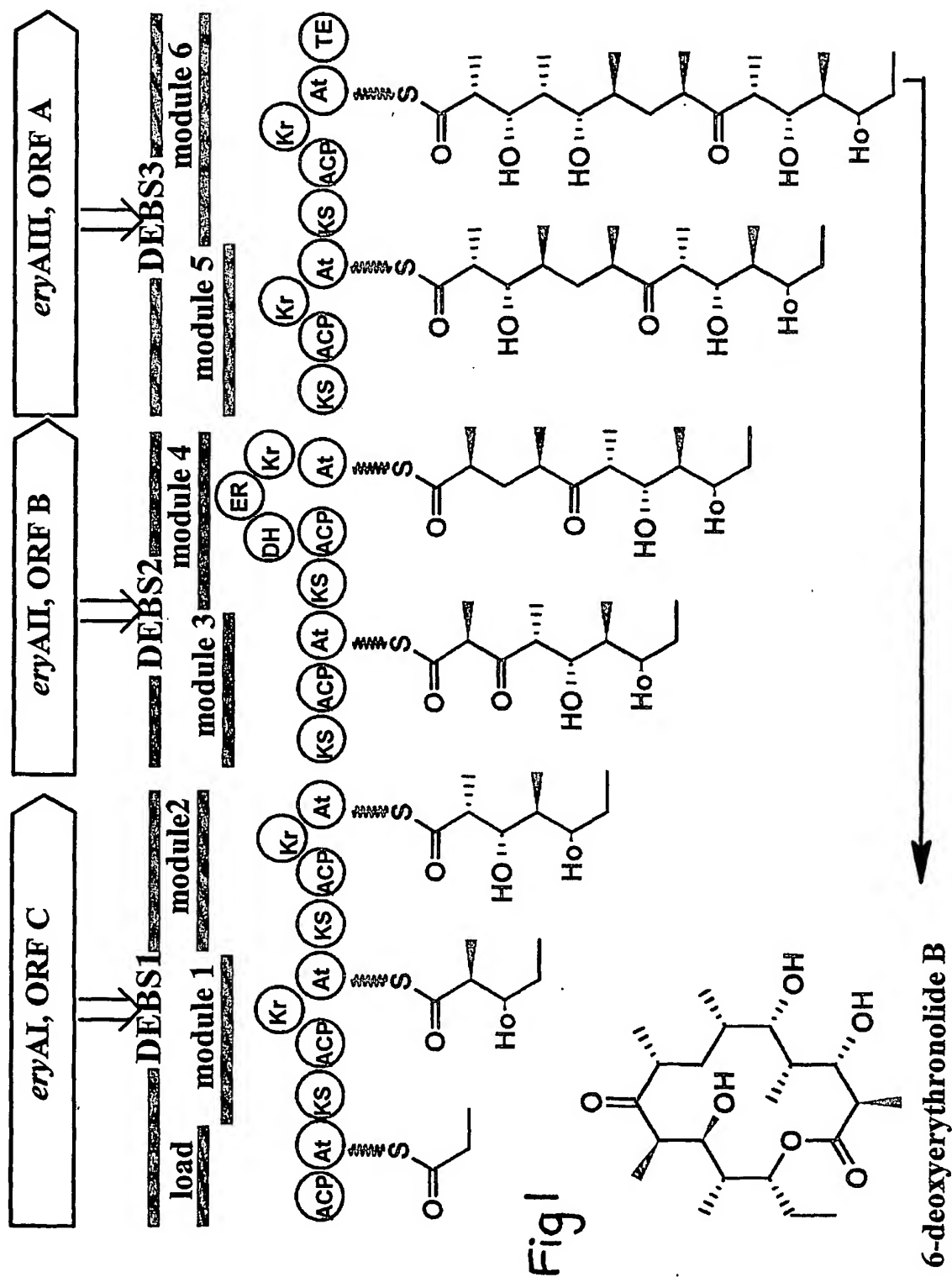
38. A method of synthesising a polyketide synthase (PKS) enzyme complex, said PKS enzyme complex including at least one acyltransferase (AT) domain; said method comprising altering said AT domain to change selectively

a minor proportion of amino acid residues, the altered residue(s) comprising one or more residues of a motif which influences the substrate specificity of the AT domain and which comprises a four-residue sequence

5 corresponding to the YASH motif of the AT domain of the first module of DEBS, the alteration affecting the substrate specificity.

39. A PKS enzyme complex as produced by the method of claims 36, 37 or 38.

1/26



	1	2/26	50
atave00x	~~~~~	~~~~~	~~~~~
atdebs00p	~~~~~	~~~~~	~~~~~
atepo06p	~~~~~	~~~~~	~~~~~
atepo07p	~~~~~	~~~~~	~~~~~
atepo01p	~~~~~	~~~~~	~~~~~
atepo05p	~~~~~	~~~~~	~~~~~
atsoralx	~~~~~	~~~~~	~~~~~
atfkb01p	~~~~~	~~~~~	~~~~~
atfkb09p	~~~~~	~~~~~	~~~~~
atrap03p	~~~~~	~~~~~	~~~~~
atrap06p	~~~~~	~~~~~	~~~~~
atrap04p	~~~~~	~~~~~	~~~~~PLVI
atrap13p	~~~~~	~~~~~A	EEAQPVETPV VASDVLPLVI
atrap01p	~~~~~	~~~~~	~~~~~
atrap07p	~~~~~	~~~~~	~~~~~PV VASELVPLVI
atrap10p	~~~~~	~~~~~	~~~~~
atfkb04x	~~~~~	~~~~~	~~~~~
attyl04p	~~~~~	~~~~~VV	REAAGRLAEV VEAGGVGLAD VAVTMAGRSR
attyl06p	~~~~~	~~~~~	~~~~~GRLAEV VEAGGVGLAD VAVTMAGRSR
attyl01p	~~~~~	~~~~~	~~~~~MAGRSR
attyl02p	~~~~~	~~~~~	~~~~~
attyl00p	~~~~~	~~~~~	~~~~~D VAVTMADRSR
atnid05b	~~~~~	~~~~~	~~~~~
attyl05b	~~~~~	~~~~~	~~~~~AAL REQSTRLLND
atnid06x	~~~~~	~~~~~	~~~~~
atdebs01p	~~~~~	~~~~~	~~~~~
atmon02p	~~~~~	~~~~~	~~~~~
atmon10p	~~~~~	~~~~~	~~~~~
atmon04p	~~~~~	~~~~~	~~~~~
atmon07p	~~~~~	~~~~~	~~~~~
atmon11p	~~~~~	~~~~~	~~~~~
atmon12p	~~~~~	~~~~~	~~~~~
atmon05b	~~~~~	~~~~~	~~~~~
atmon01p	~~~~~	~~~~~	~~~~~
atdebs02p	~~~~~	~~~~~	~~~~~
atdebs06p	~~~~~	~~~~~	~~~~~
atave01p	~~~~~	~~~~~	~~~~~
atave07p	~~~~~	~~~~~	~~~~~
atave06p	~~~~~	~~~~~	~~~~~
atave09p	~~~~~	~~~~~	~~~~~
atnys01p	~~~~~	~~~~~	~~~~~
atnys11p	~~~~~	~~~~~	~~~~~
atrif05p	~~~~~	~~~~~	~~~~~
atrif07p	~~~~~	~~~~~	~~~~~
atrif08p	~~~~~	~~~~~	~~~~~
atrif10p	~~~~~	~~~~~	~~~~~
atrif03p	~~~~~	~~~~~	~~~~~
atrif06p	~~~~~	~~~~~	~~~~~
atrif04p	~~~~~	~~~~~	~~~~~
atrif01p	~~~~~	~~~~~	~~~~~
atnys02p	~~~~~	~~~~~	~~~~~
atfkb02p	~~~~~	~~~~~	~~~~~
atavellp	~~~~~	~~~~~	~~~~~
atdebs03p	~~~~~	~~~~~	~~~~~
atnid04p	~~~~~	~~~~~	~~~~~
atdebs05p	~~~~~	~~~~~	~~~~~
atdebs04p	~~~~~	~~~~~	~~~~~

Fig 2a

3/26

atave02a	LFAFQVALHR	LLTDGYHITP	HYYAGHSLGE	ITAAHLAGIL	TLTDATTLIT
atave05a	LFAFQVALHR	LLTDGYHITP	HYYAGHSLGE	ITAAHLAGIL	TLTDATTLIT
atave04a	LFAFQVALHR	LLTDGYHITP	HYYAGHSLGE	ITAAHLAGIL	TLTDATTLIT
atave08a	LFAFQVALHR	LLTDGYHITP	HYYAGHSLGE	ITAAHLAGIL	TLTDATTLIT
atave03a	LFAFQVALHR	LLTDGYHITP	HYYAGHSLGE	ITAAHLAGIL	TLTDATTLIT
atrap02a	LFALQVALFG	LL.ESWGVPR	DAVVGHSVGE	LAAGYVSGLW	SLEDACLTVS
atrap11a	LFALQVALFG	LL.ESWGVPR	DAVVGHSVGE	LAAAYVSGVW	SLEDACLTVS
atrap08a	LFALQVALFG	LL.ESWGVPR	DAVVGHSVGE	LAAGYVSGLW	SLEDACLTVS
atrap12a	LFAMQVALFG	LL.ESWGVPR	DAVVGHSVGE	LAAAYVSGVW	SLEDACLTVS
atrap05a	LFALQVALFG	LL.ESWGVPR	DAVVGHSVGE	LAAGYVSGLW	SLEDACLTVS
atrap09a	LFALQVALFG	LL.ESWGVPR	DAVVGHSVGE	LAAAYVSGLW	SLEDACLTVS
atfkb03a	VFALQVALSA	QL.DAWGVPR	DVLVGHSGE	LAAAYVAGVW	SLDDATELVS
atfkb07x	HFAHQIALTA	LL.RSWGITP	HAVIGHSLGE	ISAACAAGVL	SIGDASALLA
atfkb08x	LFAHQAAFTA	LL.RSWDITP	HAVIGHSLGE	ITAAYAAGIL	SLDDACTLIT
atnid01a	LFALQTALYR	TL.TARGETA	HLVLGHSGE	ITAAHIAGVL	DLPDAARLIT
atnid03a	LFALQTALYR	TL.TARGETA	HLVLGHSGE	ITAAHIAGVL	DLPDAARLIT
atnid02a	LFALQTALYR	TL.TARGETA	HLVLGHSGE	ITAAHIAGVL	DLPDAARLIT
atnid00a	LFALQTALYR	TL.TARGETA	HLVLGHSGE	ITAAHIAGVL	DLPDAARLIT
atfkb10a	LFTLEVALLR	LL.EHWGVPR	DVVVGHSVGE	VTAAYAAGVL	TLADATTLIV
atrap14a	IFAMEAALFG	LL.EDWGVPR	DFVAGHSIGE	ATAAYASGML	SLNVTTTLIV
atmon06a	LFALQVGLAR	LW.ESVGVPR	DVVLGHSGE	IAAAHVAGVF	DLADACRVVG
atmon08a	LFALQVGLAR	LW.ESVGVPR	DVVLGHSGE	IAAAHVAGVF	DLADACRVVG
atmon09a	LFALQVGLAR	LW.ESVGVPR	DVVLGHSGE	IAAAHVAGVF	DLADACRVVG
atepo02a	LFAVEYALTA	LW.RSWGVEP	ELLYGHSGE	LVAACVAGVF	SLEDGVRIVA
atepo03x	LFTVEYALTA	LW.RSWGVEP	ELVAGHSAGE	LVAACVAGVF	SLEDGVRIVA
atepo08a	LFALEYALAA	LF.RSWGVEP	ELVAGHSLGE	LVAACVAGVF	SLEDAVRLVV
atepo00a	LFTFEYALAA	LW.RSWGVEP	ELVAGHSIGE	LVAACVAGVF	SLEDAVRLVA
atepo04a	LFALEYALAA	LW.RSWGVEP	HVLLGHSGE	LVAACVAGVF	SLEDAVRLVA
atnid07a	LFAVETALFR	LF.ESWGLMP	DVLLGHSGIG	LAAAYAGVF	SSADAVRLVA
attyl07a	LFAVEVALHR	LL.EHWGMRP	DLLLGHSGE	LAAAHVAGVL	DLDDACALVA
atsor02a	LFALEVALFQ	LL.QSFGCLKP	ALLLGHSIGE	LVAAHVAGVL	SLQDACLIVA
atsorb1a	LFALEVALFE	LL.QSFGCLKP	ALLLGHSIGE	LVAAHVAGVL	SLQDACLIVA
atnys09a	LFAVEVALYR	LI.ESFGVPR	DHLAGHSVGE	IVAHLAGVL	SLADAATLVA
atnys12a	LFAVEVALFR	LL.TSWGLTP	DYLAGHSVGE	LAAAHVAGVL	SLDDACLIVA
atnys16a	LFAVEVALFR	LV.ASWGVP	EFVAGHSVGE	IAAAHVAGVF	SLVDACRLVV
atnys17a	LFAVEVALFR	LV.ASWGVP	EFVAGHSVGE	IAAAHVAGVF	SLVDACRLVV
atnys03a	LFAVEVALYR	LV.ASLGVTP	DFVGGHSIGE	LAAAHVAGVL	SLEDACLIVA
atnys15a	LFAVEVALYR	LI.ESWGVAP	DFVAGHSIGE	IAAAHVAGVF	SLEDACLIVA
atnys07a	LFAIEVALFR	LV.ESWGVPR	DFVAGHSIGE	IAAAHVAGVF	SLEDACRLVA
atnys08a	LFAVEVALFR	LV.ERWGVPR	DFVAGHSIGE	IAAAHVAGVF	SLEDACRLVA
atnys05a	LFAVEVALFR	LV.ESWGVPR	DFVAGHSIGE	IAAAHVAGVF	SLEDACRLVA
atnys06a	LFAIEVALFR	LV.ESWGVPR	DFVAGHSIGE	IAAAHVAGVF	SLEDACRLVA
atnys04a	LFAIEVALFR	LL.EAWGITP	DFVAGHSIGE	IAAAHVAGVL	SLGDACRLVV
atnys14a	LFAVEVALYR	LI.ESWGVPR	DFVAGHSVGE	LAAAHVAGVL	SLDDACRLVA
atnys00a	LFAVEVALHR	LV.ASLGVTP	DFVGGHSVGE	IAAAHVAGVL	SLEDACRLVA
atnys10a	LFAVEVALFR	LV.ESWGVPR	DFVAGHSIGE	IAAAHVAGVL	TLEDACRLVA
atnys18a	LFAVEVALYR	LI.ASWGIRP	DHVTGHSIGE	ITAAHVAGVL	TLADACLIVA
atnys13a	LFAVEVALFR	LA.ESWRLTP	DFVAGHSIGE	IAAAHVAGVF	SLEDACLIVA
atavel0a	LFAFEVALFR	LL.ETWGLTP	DVVLGHSGE	LAAAHVAGML	CLADAVALVV
atrif02a	LFAVETALFR	LF.ESWGVPR	GLLAGHSIGE	LAAAHVSGVL	DLADAGELVA
atmon03a	LFALEVALYR	QV.TSFGIAP	SHLTGHSVGE	IAAAHVAGVF	SLADACLIVA
atavel2a	LFAVQVALFR	HL.ERLGVR	DFVAGHSIGE	LAAAHVAGVL	PLAAACRLVA
atrif09a	LFAVESALFR	LA.ESWGVPR	DVVLGHSGE	ITAAYAAGVF	SLPDARIVA
atmon00a	LFAIETSLYR	LA.ASFGLKP	DVVLGHSGE	IAAAHVAGVL	SLPDASALVA
attyl03a	LFALQTALFR	LA.EHHGLRA	EALCGHSVGE	IAAAHAAGVL	TLPDAARLVA

GHS

Fig2j

4/26

251	atave00x LWSQAQTT.L AGTGALVSVA ATPDELLPRI APWTEEN.PA RLAVAAVNGP atdebs00p LWSREMIP.L VGNGDMAAVA LSADEIEPRI ARWDDD.... .VVLAVNGP atepo06p RRSRL.L.RRI SGQGEMALVE LSLEEAEALRGHEG RLSVAVSNSP atepo07p RRSRL.L.RRI SGQGEMALVE LSLEEAEALRGHEG RLSVAVSNSP atepo01p RRSRL.L.RRI SGQGEMAVTE LSLAEAEALRGYED RVSVAVNSP atepo05p RRSLLL.L.RRI SGQGEMAVVE LSLAEAEALLGYED RLSVAVSNSP atsorax AYGRII.RKL RGKGGMLVA LSWEDAGKELTGYEG RLFRAIEHSA atfkb01p LRSQAIAARL AGRGAMASIA VPASAVE...TVE GVWIAARNGP atfkb09p LRSQTIAAHL AGRGAMASIA LPATAVE...TVE GVWVAARNGP atrap03p LRSQAIAARL AGRGAMASIA LPAQDVE...LVD GAWIAAHNGP atrap06p LRSQAIAARL AGRGAMASIA LPAQDVE...LVD GAWIAAHNGP atrap04p LRSQAIAARL AGRGAMASIA LPAQDVE...LVD GAWIAAHNGP atrap13p LRSQAIAARL AGRGAMASIA LPAQDVE...LVD GAWIAAHNGP atrap01p LRSQAIAARL AGRGAMASIA LPAQDVE...LVD GAWIAAHNGP atrap07p LRSQAIAARL AGRGAMASIA LPAQDVE...LVD GAWIAAHNGP atrap10p LRSQAIAARL AGRGAMASIA LPAQDVE...LVD GAWIAAHNGP atfkb04x LRSALLVREL AGRGAMASIA FAA..AA...RID GVWVAGRNGT atty104p LRAGLIGRYL AGRGAMAAPV LPAGEVEAGL AKWPG..... .VEVAAVNGP atty106p LRAGLIGRYL AGRGAMAAPV LPAGEVEAGL AKWPG..... .VEVAAVNGP atty101p LRAGLIGRYL AGRGAMAAPV LPAGEVEAGL AKWPG..... .VEVAAVNGP atty102p LRAGLIGRYL AGRGAMAAPV LPAGEVEAGL AKWPG..... .VEVAAVNGP atty100p LRAGLIGRYL AGRGAMAAPV LPAGEVEAGL AKWPG..... .VEVAAVNGP atnid05b LRSRAWLG.L AGKGGMVAVP MPAEELRPLVTWGD RLAVAAVNSP atty105b LRSRAWLT.L AGKGGMAAVS LPEARLRERIERFQG RLSVAAVNSP atnid06x GRSRLWGR.L AGNGGMLAVM APAERIRELLEPWRQ RISVAAVNGP atdebs01p LRSRVIAAT.M PGNGGMAASIA APAGEVRARIGD RVEIAAVNGP atmon02p VRSDAL.RQL QGHGDMASLS TGAEQAAELI GDRPG..... .VVVAAVNGP atmon10p VRSDAL.RQL QGHGDMASLS TGAEQAAELI GDRPG..... .VVVAAVNGP atmon04p VRSDAL.RQL QGHGDMASLG TGAEQAAELI GDRPG..... .VVVAAVNGP atmon07p VRSDAL.RQL MGQGDMSLG ASSEQAAELI GDRPG..... .VCIAAVNGP atmon11p VRSDAL.RQL QGHGDMASLS TGAEQAAELI GDRPG..... .VVVAAVNGP atmon12p VRSDAL.RQL MGQGDMSLG AGSEQVAELI GDRPG..... .VCVAAVNGP atmon05b VRSVLL.RQL SGRGGMASLG MGQEQAADLI DGHPG..... .VVVAAVNGP atmon01p LRSRAL.RQL SGGGAMASLG VGQEQAAELV EGHPG..... .VGIAAVNGP atdebs02p RRSRAV.RAV AGRGMSLSVR GGRSDVEKLL ADDS...WTG RLEVAAVNGP atdebs06p LRAKAL.RAL AGKGGMVSLA APGERARALI A..P...WED RISVAAVNSP atave01p LRSRALAA.V RGRGGMASVP LPAQEVEQLIGERWAG RLWVAAVNGP atave07p LRSRALAA.V RGRGGMASVP LPAQEVEQLIGERWAG RLWVAAVNGP atave06p LRSRALAA.V RGRGGMASLP LPAQDVQQLISERWEG QLWVAALNGP atave09p LRSQALAA.V RGRGAMVSLP LPAQDVQQLISERWEG QLWVAALNGP atnys01p LRSQAIGRAL AGRGGMSVA LSVDVLEPRL VE.....FEG RVSVAAVNGP atnys11p LRSQAIGRAL AGRGGMSVA LSVDVLEPRL VE.....FEG RVSVAAVNGP atrif05p LRSQAIAAEL SGRGGMASIQ LSHDEVAARL AP.....WAG RVEIAAVNGP atrif07p LRSQAIAARL SGRGGMASVA LSEDEANARL GL.....WDG RIEVAAVNGP atrif08p LRSQAIAAKL AGRGGMASVA LSEEDAVARL RH.....WAD RVEVAAVNSP atrif10p LRSQAIAAKL SGRGGMASVA LGEADVVSRLAD GVEVAAVNGP atrif03p LRSQAIAAGL AGRGGMASVA LSEEDAVARL TP.....WAN RVEVAAVNSP atrif06p LRSQAIAATRL AGRGGMASVA LSEEDATAWL AP.....WAD RVQVAAVNSP atrif04p LRSQAIAASL AGRGGMASVA LSEEDATARL EP.....WAG RVEVAAVNGP atrif01p LRSQAIAAEL SGRGGMASVA LGEDDVVSRLVD GVEVAAVNGP atnys02p LRSQALP.QL SGRGGMSVS APVERVTALL AP.....WQE ALSVAAVNGP atfkb02p LRSRLVATER AGHGGMVSV PAFDAAA...WAG RLEVAAVNGP atave11p LRSQALA.AL AGQGAMASVG LPVEKLEPRL A.....TWGD RLVAIAAVNGA atdebs03p GRSRLM.RSL SGEGMAAVA LGAAVRERL RPWQ.....D RLSVAAVNGP atnid04p LRSQLIAREL AGRGMSASVA LAAADVESRL AGAEAGGGVR DVEIAAVNGP atdebs05p VRSRVL.RRL GGQGGMASFG LGTEQAAERIGRFAG ALSIASVNGP atdebs04p LRSQVL.REL DDQGGMVSVG ASRDELETVL A.....RWDG RVAVAAVNGP	300
-----	---	-----

Fig 2k

5/26

atave02a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atave05a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atave04a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atave08a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atave03a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atrap02a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atrap11a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atrap08a	~~~~~	~~~~~	~~~~~	~PPTQPADNA	VIERAPEWLP
atrap12a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atrap05a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atrap09a	~~~~~	~~~~~	~~~~~	~D	DVRPADAPVV ASVMASELVP
atfkb03a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atfkb07x	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atfkb08x	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnid01a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnid03a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnid02a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnid00a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atfkb10a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atrap14a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atmon06a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atmon08a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atmon09a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atepo02a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atepo03x	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atepo08a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atepo00a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atepo04a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnid07a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atty107a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atsor02a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atsorb1a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys09a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys12a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys16a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys17a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys03a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys15a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys07a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys08a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys05a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys06a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys04a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys14a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys00a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys10a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys18a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys13a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atave10a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atrif02a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atmon03a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atave12a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atrif09a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atmon00a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atty103a	SVPAGEPPAA	GRPEDTGGAW	TVSGRGPAAL	RAQAARLYDA	LTGTGTGTGQ

Fig 2b

6/26

	51				100
atave00x	~~~~~	~~~~~	~~~~~	~~~~~VQR	MDGGEEPRA
atdebs00p	~~~~~	~~~~~	~~~~~	~~~~~	~~~VADGRPH
atepo06p	~~~~~	~~~~~	~~~~~	~~~~~	~~~AAAQGHTP
atepo07p	~~~~~	~~~~~	~~~~~	~~~SSREALGA	LSAAAQGHTP
atepo01p	~~~~~	~~~~~	~~~~~	~~~~~REG	LDAAARGQTP
atepo05p	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~P
atsoralx	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atfkb01p	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atfkb09p	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atrap03p	SAKTQPALTE	HEDRLRAYLA	ASPGADTRAV	ASTLAVTRSV	FEHRAVLLGD
atrap06p	~~~TQPALTE	HEDRLRAYLA	ASPGVDTRAV	ASTLAVTRSV	FEHRAVLLGD
atrap04p	SAKTQPALTE	HEDRLRAYLA	ASPGADTRAV	ASTLAVTRSV	FEHRAVLLGD
atrap13p	SAKTQPALTE	HEDRLRAYLA	ASPGADIRAV	ASTLAVTRSV	FEHRAVLLGD
atrap01p	~~~~~	~~~~~	~~~~~	~~~LAVTRSL	FEHRAVLLGD
atrap07p	SAKTLPALTE	HEDRLRAYLA	ASPGADMRAV	GSTLALTRSV	FEHRAVLLGH
atrap10p	~~~~~	~~~~~	~~~~~AV	ASTLAVTRSV	FEHRAVLLGD
atfkb04x	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atty104p	FGYRAVVLAR	GEAELAGRLR	ALAGGDPDAG	VVTGAVVD..
atty106p	FGYRAVVLAR	GEAELAGRLR	ALAGGDPDAG	VVTGAVVD..
atty101p	FGYRAVVLAR	GEAELAGRLR	ALAGGDPDAG	VVTGAVVD..
atty102p	~~~~~	~~~~~RLR	ALAGGDPDAG	VVTGAVVD..
atty100p	FGYRAVVLAR	GEAELAGRLR	ALAGGDPDAG	VVTGAVLDGG	VVVGAAAPGGA
atnid05b	~~~~~	~~~~~LLSTR	ARFPRAAVV	GESMTELAEA	LDAAVEGGPH
atty105b	LLEHPDEHPA	DVGYTITGR	AHFGHRAAVI	GESREELDA	LKALAEGRH
atnid06x	~~~~~	~~~~~	~~~~~	~~~~~	~RSVAERPE
atdebs01p	~~~~~	~~~~~	~~~~~	~~~~~	~~GLATGNAD
atmon02p	~~~~~	~~~~~	~~~~~	~~~~~	~GALAAGEAS
atmon10p	~~~~~	~~~~~	~~~~~	~~~~~	LGALAAGEAS
atmon04p	~~~~~	~~~~~	~~~~~	~~~~~	~~~LAAGETP
atmon07p	~~~~~	~~~~~	~~~~~	~~~~~	~~ALAAGEES
atmon11p	~~~~~	~~~~~	~~~~~	~~~~~	~~ALAAGEAS
atmon12p	~~~~~	~~~~~	~~~~~	~~~~~	~~~LAAGEPS
atmon05b	~~~~~	~~~~~	~~~~~	~~~~~	~~~SLAAGEAS
atmon01p	~~~~~	~~~~~	~~~~~	~~~~~	~EALAAGDAS
atdebs02p	~~~~~	~~~~~	~~~~~	~~~~~	~~~ADGAVV
atdebs06p	~~~~~	~~~~~	~~~~~	~~~~~	~RAVAEGVAA
atave01p	~~~~~	~~~~~	~~~~~	~~~~~G	LGALAAGEPD
atave07p	~~~~~	~~~~~	~~~~~	~~~~~G	LGALAAGEPD
atave06p	~~~~~	~~~~~	~~~~~	~~~~~QA	LTALAAGEPH
atave09p	~~~~~	~~~~~	~~~~~	~~~~~	LTALAAGEPH
atnys01p	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys11p	~~~~~	~~~~~	~~~~~	~~~~~	~~~AVATDG
atrif05p	~~~~~	~~~~~	~~~~~	~~~~~	~TALARGESA
atrif07p	~~~~~	~~~~~	~~~~~	~~~~~G	LGALARGEAA
atrif08p	~~~~~	~~~~~	~~~~~	~~~~~AG	LAALARGESA
atrif10p	~~~~~	~~~~~	~~~~~	ADSAAEEARAG	LGALARGEDA
atrif03p	~~~~~	~~~~~	~~~~~	~~~~~QDG	LQALARGENA
atrif06p	~~~~~	~~~~~	~~~~~	~~~SREEAVTN	LEALARGEDP
atrif04p	~~~~~	~~~~~	~~~~~	~~~~~	~RALARGESA
atrif01p	~~~~~	~~~~~	~~~~~	~~~~~V	VVAGSREEAV
atnys02p	~~~~~	~~~~~	~~~~~AVVV	GERREDFLRG	LAALSTGAST
atfkb02p	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~GEEV
atavellp	~~~~~	~~~~~	~~~~~	~~~~~LHA	LDALAEGAPT
atdebs03p	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~AATA
atnid04p	~~~~~SLADS	AGIGHGLAVG	RAALPHRAVL	LGDGAAPLDA	LAALASGEVS
atdebs05p	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~ADRRRIA
atdebs04p	~~~~~	~~~~~	~~~~~	~~~~~	~~~ALAEGRPS

Fig 2c

7/26

atave02a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atave05a	~~~~~	~~~~~	~~~~~	~~~~~QALQAL	AAGEPHEPAVI
atave04a	~~~~~	~~~~~	~~~~~	~~~~~QALQAL	AAGEPHEPAVI
atave08a	~~~~~	~~~~~	~~~~~	~~~~~QALQAL	AAGEPHEPAVI
atave03a	~~~~~	~~~~~	~~~~~	~~~~~QALQAL	AAGEPHEPAVI
atrap02a	~~~~~	~~~~~	~~~~~DT	RAVASTLAMT	RSVFYRAVL
atrap11a	~~~~~	~~~~~	~~~~~	~AVASTLAMT	RSMFEHRGVL
atrap08a	MVISARTQSA	LTEHEGRLRA	YLAASPGVDM	RAVASTLAIT	RSVFHRAVL
atrap12a	~~~~~	LTEHEGRLRA	YLAASPGVDM	RAVASTLAMT	RSVFHRAVL
atrap05a	~~~~~	~~~~~	~~~~~	~~~ASTLAVT	RSVFHRAVL
atrap09a	LVISAKTQSA	LAIEYEGRLRA	YLAASPGVDM	RAVASTLAMT	RSVFHRAVI
atfkb03a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~HRAAL
atfkb07x	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~L
atfkb08x	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnid01a	~~~~~	~~~~~	~~~~~KHRA	VITGRTRTEL	HTKLHTLDAI
atnid03a	~~~~~TQA	DPQDIAHALA	TTRTHFKHRA	VITGRTRTEL	HTKLHTLDAI
atnid02a	~~~~~	~~~~~HALA	TTCTHFKHRA	VITGRTRTEL	HTKLHTLDAI
atnid00a	~~~~~	~~~~~	~~~~~	~~~~~	SSALAALAAG
atfkb10a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atrap14a	~~~~~	~~~~~	~~~~~	~~~~~DFLRA	LSKLADGAPW
atmon06a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~TGEPHA
atmon08a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~AGEEHP
atmon09a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~GEEHP
atepo02a	~~~~~	~~~~~	~~~~~	~~~~~	AALSAVAQGQ
atepo03x	~~~~~	~~~~~	~~~~~	~~~~~A	VAVTSREGLL
atepo08a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~VAAQGQ
atepo00a	~~~~~	~~~~~	~~~~~	~~~~~SREGLR	AALDAAAQGG
atepo04a	~~~~~	~~~~~	~~~~~	~~~~~LR	GALDAAAQOK
atnid07a	~~~~~	~~~~~	~~~~~A	AAHDALLAVA	DGRPSDAVVT
atty107a	PRDIAFSLAA	TRAAFDHRAV	LIGSDGAELA	AALDAL...A	EGRDGPVVR
atsor02a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atsorbla	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
atnys09a	~~~~~	~~~~~	~~~~~	~~~~~AD	DPAAAPAWIT
atnys12a	~~~~~	~~~~~	~~~~~	~~~~~S	DGRPDGVLVQ
atnys16a	~~~~~	~~~~~	~~~~~	~~~~~	~PD.LPEVAR
atnys17a	~~~~~	~~~~~	~~~~~	~~~~~	APDGITAAAR
atnys03a	~~~~~	~~~~~	~~~~~GYALADGR	ATFEHRAVLL	PDGTELA..H
atnys15a	~~~~~	~~~~~	~~~~~	~~~~~	PDAHE.G..H
atnys07a	~~~~~	~~~~~	~~~~~	~~~~~IAA	DEA.DAAAAT
atnys08a	~~~~~	~~~~~	~~~~~	~~~~~ALAALAS	GVA.DPAVVS
atnys05a	~~~~~	~~~~~	~~~~~	~~~~~	AVRALTALAA
atnys06a	~~~~~	~~~~~	~~~~~	~~~~~	ATRALSALAT
atnys04a	~~~~~	~~~~~HR	AVVLGTDRAE	ALRALALAA	GE.TDPAALT
atnys14a	~~~~~	~~~~~	~~~~~DG	LRTGLTAVAE	GTTAPHTAEH
atnys00a	~~~~~	~~~~~	~~~~~	~~~~~	~ADAVEHAR
atnys10a	~~~~~	~~~~~	~~~~~VVAQDRDQ	LIASLGALAA	DRPDPAVVEG
atnys18a	~~~~~	~~~~~	~~~~~	~~~~~	EGGAVTEVAR
atnys13a	~~~~~	~~~~~	~~~~~	~~~~~LLA	GPDGVREAR
atave10a	~~~~~	~~~~~	~~~~~	~~~~~LHALDALA	GGRPVPGVVE
atrif02a	~~~~~	~~~~~	~~~~~R	AVVLASDRAQ	LCADLAAFGS
atmon03a	~~~~~	~~~~~	~~~~~	~~~~~A	LAAGRAHPAL
atave12a	~~~~~	~~~~~	~~~~~	~~~~~QALDALA	EGRSADGLIE
atrif09a	~~~~~	~~~~~	~~~~~GRALLGDR	AVVAGTDED	AVAGLRALAR
atmon00a	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~LAEG
atty103a	GAGQGAGPGT	AEVAGALAHA	RTAFRHRAV	LGGNRAELLA	GLRELAEEHH

Fig 2d

8/26

	101				150
atave00x	AGEVLGVADE	ADGG..VVFV	FPGQGPQWPG	MGRELLDASD	VFRESVRACE
atdebs00p	ASVVRGVA.R	PSAP..VVFV	FPGQGAQWAG	MAGELLGESR	VFAAAMDACA
atepo06p	PGAVRGRASG	GSAP.KVVFV	FPGQGSQWVG	MGRKLM AE EP	VFRAALEGCD
atepo07p	PGAVRGRASG	GSAP.KVVFV	FPGQGSQWVG	MGRKLM AE EP	VFRAALEGCD
atepo01p	PGAVRGRCSF	GNVP.KVVFV	FPGQGSQWVG	MGRQLLAE EP	VFHAALSACD
atepo05p	PAAARGHAST	GSAP.KVVFV	FPGQGSQWLG	MGQKLLSE EP	VFRDALSACD
atsoralx	~~~~~	~~~~~VVFV	FAGQGAQWFG	MGRALLQREP	VFRTTIEQCS
atfkb01p	~~~~SAVAGV	AVEGARTVFV	FPGQGSQWVG	MGRELMGASE	VFAARMRECA
atfkb09p	~~~~~	~~~~~VVFV	FPGQGSQWVG	MGRELMGCSE	VFAARMRECA
atrap03p	D..TV..TGT	AVSDPRVVFV	FPGQGWQWLG	MGSALRDSSV	VFAERMAECA
atrap06p	D..TV..TGT	AVSDPRVVFV	FPGQGWQWLG	MGSALRDSSI	VFAERMAECA
atrap04p	D..AV..TGT	AVTDP RVVFV	FPGQGWQWLG	MGSALRDSSV	VFAERMAECA
atrap13p	D..TV..TGT	AVTDP RVVFV	FPGQGWQWLG	MGSALRDSSV	VFAERMAECA
atrap01p	D..SVTGTGT	AVSDPRVVFV	FPGQGWQWLG	MGSALRTSSM	VFAERMAECA
atrap07p	DTVTVTGTGT	AVSNPRVVFV	FPGQGWQWLG	MGSALRGSSV	VFAERMAECA
atrap10p	ETV....TGT	AVSDPRVVFV	FPGQGWQWLG	MGSALRDSSV	VFAERMAECA
atfkb04x	~~~~VVTGT	ALTAPRTVFV	FPGQGSQWLG	MGRELM AE SP	VFAARMRQCA
attyl04pPET	GSGGGGVVLV	FPGQGTQWVG	MGAGLLGSSE	VFAASMRECA
attyl06pPET	GSGGGGVVLV	FPGQGTQWVG	MGAGLLGSSE	VFAASMRECA
attyl01pPET	GSGGGGVVLV	FPGQGTQWVG	MGAGLLGSSE	VFAASMRECA
attyl02pPET	GSGGGGVVLV	FPGQGTQWVG	MGAGLLGSSE	VFAASMRECA
attyl00p	GAAGGAGAAG	GAGGGGVVLV	FPGQGTQWVG	MGAGLLGSSE	VFAASMRECA
atnid05b	..PLAATGT.	AGTADRVVFV	FPGQGSQWAG	MAEGLLERSG	AFRSAADSCD
attyl05b	HTVVRGDGT.	AHPDRRVVFV	FPGQGSQWPS	MARDLLDRAP	AFRETAACD
atnid06x	PDVVL..GE.	AGSDRAPAFV	FPGQGAQWAG	LGARLLADSP	VFRARA EACA
atdebs01p	GA AV...GT.	SRAQQR AVFV	FPGQGWQWAG	MAVDLLDTSP	VFAAALRECA
atmon02p	AGVVAG.VAG	DVGPGP.VLV	FPGQGAQWVG	MGAQLLDESA	VFAARIAECE
atmon10p	AGVVAG.VAG	DVGPGP.VLV	FPGQGSQWVG	MGAQLLDESP	VFAARIAECE
atmon04p	TDVVSG.AAA	SSGAGP.VLV	FPGQGSQWVG	MGAQLLDESP	VFAARIAECE
atmon07p	ADV VAG.VAG	DVGPGP.VLV	FPGQGSQWVG	MGAQLLDESP	VFAARIAECE
atmon11p	ADV VAG.VAG	DVGPGP.VLV	FPGQGSQWVG	MGAQLLDESP	VFAARIAECE
atmon12p	PDVVEGAVQG	ASGAGP.VLV	FPGQGSQWVG	MGAQLLDESP	VFAARIAECE
atmon05b	PDVVS GAV.G	PTGPGP.VMV	FPGQGGQWVG	MGARLLDESP	VFAARIAECE
atmon01p	PDV VCG.VAG	DVGPGP.VLV	FPGQGSQWVG	MGAQLLGESA	VFAARIDACE
atdebs02p	PGVVTGSASDGGSVFV	FPGQGAQWEG	MARELL.PVP	VFAESIAECD
atdebs06p	PGATTGTASAGGVVFV	FPGQGAQWEG	MARGLL.SVP	VFAESIAECD
atave01p	RRVTTGHAPG	GDRGG.VVFV	FPGQGGQWAG	MGVRL LASSP	VFARRMQACE
atave07p	RRVTTGHAPG	GDRGG.VVFV	FPGQGGQWAG	MGVRL LASSP	VFARRMQACE
atave06p	PHITTGHTRG	GDRGG.VVFV	FPGQGGQWAG	MGLTLLTSSP	VFAEHIDACE
atave09p	PHITTGHTRG	SDRGG.VVFV	FPGQGGQWAG	MGLTLLTSSP	VFAEHIDACE
atnys01p	~~~~~L.	ADVEGRTVFV	FPGQGSQWVG	MGAQLLDESA	VFAERIAECA
atnys11p	PSPVVARGV.	ADVEGRTVFV	FPGQGSQWVG	MGSQLLDESA	VFAERIAECA
atrif05p	SGLVTGT...	AGMPGKTVVW	FPGQGTQWAG	MGRELLEASP	VFAERIEECA
atrif07p	PGVVTGT...	AGKPGKVVVW	FPGQGTQWVG	MGRELLDASP	VFAERIKECA
atrif08p	ADVVTGTVAA	SGVPGKLVVW	FPGQGSQWVG	MGRELLEASP	VFAARIAECA
atrif10p	PGLVRGRVPA	SGLPGKLVVW	FPGQGTQWVG	MGRELLEESP	VFAERIAECA
atrif03p	PGVVTGT...	AGKPGKVVVW	FPGQGSQWVG	MGRDLLDSSP	VFAARIKECA
atrif06p	AAVVTGR...	AGSPGKLVVW	FPGQGSQWIG	MGRELLDSSP	VFAERVAECA
atrif04p	PGLLSGR..G	SGVPGKVVVW	FPGQGTQWAG	MGRELLDSSE	VFAARIAECE
atrif01p	TGLRALNTAG	SGTPGKVVVW	FPGQGTQWAG	MGRELLAE SP	VFAERIAECA
atnys02p	AGLVSG..IA	GPDPEGAVFV	FPGQGSQWVG	MGRELLATSE	VFR TAID DCA
atfkb02p	PGVVRGTADV	TDT..RAVFV	FPGQGSQWDG	MGAELLATEP	VFARRLG ECA
atave11p	AGVVQGVAGP	AA.DGKIAML	FPGQGTQWVG	MAQELLGSSP	VFAQQMSDCA
atdebs03p	DAVVEGV.TE	VD.GRNVVFL	FPGQGSQWAG	MGAELLSSSP	VFAGKIRACD
atnid04p	PDVVTG..SA	AD.VRRVAFV	FPGQGAQWAG	MGAELLDSSP	VFAAELARCE
atdebs05p	DRTATGQ.GP	NS.PRRVAMV	FPGQGAQWQG	MARDLLRESQ	VFADSI RDCE
atdebs04p	ADAVAPV TSA	...PRKPV LV	FPGQGAQWVG	MARDLLESSE	VFAESMSRCA

Fig 2e

9/26

```

atave02a HSSAPGGTGT GEAAGKTAFI CSGQGTQRPQ MAHGLYHTHP VFAAALNDIC
atave05a HSSAPGGTGT GEAAGKTAFI CSGQGTQRPQ MAHGLYHTHP VFAAALNDIC
atave04a HSSAPGGTGT GEAAGKTAFI CSGQGTQRPQ MAHGLYHTHP VFAAALNDIC
atave08a HSSAPGGTGT GEAAGKTAFI CSGQGTQRPQ MAHGLYHTHP VFAAALNDIC
atave03a HSSAPGGTGT GEAAGKTAFI CSGQGTQRPQ MAHGLYHTHP VFAAALNDIC
atrap02a IGDDTVTG.T AATDPRVVFV FPGQGSQRAG MGEELAAAFP VFARIHQQVW
atrap11a LGDGTVSG.T AVSDPRVVFV FPGQGSQRAG MGEELAAAFP VFARIHQQVW
atrap08a LGDDTVTG.T AATDPRVVFV FPGQGSQRAG MGEELAAAFP VFARIHQQVW
atrap12a LGDDTVTG.T AVSDPRVVFV FPGQGSQRAG MGEELAAAFP VFARIHQQVW
atrap05a LGDDTVTG.T TVSDPRVVFV FPGQGSQRAG MGEELAAAFP VFARIHQQVW
atrap09a VGDDTVSG.T AATDPRVVFV FPGQGSQRAG MGEELAAAFP VFARIHQQVW
atfkb03a IGTDLTITG.T AEPDRRLVWL FSGQGSQRPG MGDELAAYD VFARTRRDVL
atfkb07x LGDTLITADP NAGSGPVVFV YSGQSTLHPH TGHQLAATYS VFADAWGEVL
atfkb08x ~~~~~IGAPP ADQADELVFV YSGQGTQHPA MGEELAAAFP VFADAWHDAL
atnid01a Q.....GT AHPHPRLTLL FTGQGAQHPRG MGQELYATDP HFAAALDEVC
atnid03a Q.....GT AHPHPRLTLL FTGQGAQHPRG MGQELYTTDP HFAAALDEVC
atnid02a Q.....GT AHPHPRLTLL FTGQGAQHPRG MGQELYTTDP HFAAALDEIC
atnid00a QTPRGVRIGS TDADGRLALL FTGQGAQHPRG MGQELYTTDP HFAAALDEVC
atfkb10a ~~~EAPESSA EPPRSARRFL FDGQGAQRVG MGRELHGRFP VFAAAWDEVS
atrap14a PGLTTATATA KARRVA..FL FDGQGTQRLG MGKELYDSYP AFARAWDTVS
atmon06a ALVGPAACSQA RVGGDDVWVL FSGQGSQVLG MGAGLYERFP VFAAAFDEVC
atmon08a AVTRSREDGV AASG.AVVWL FSGQGSQVLG MGAGLYERFP VFAAAFDEVC
atmon09a AVTRSREEAA VAASGDVWVL FSGQGSQVLG MGAGLYERFP VFAAAFDEVC
atepo02a TPAGAARCIA SSSRGKLAFL FTGQGAQTPG MGRGLCAAWP AFREAFDRCV
atepo03x TPPGAARCIA SSSRGKLAFL FTGQGAQTPG MGRGLCAAWP AFREAFDRCV
atepo08a TPAGAARGRA ASSPGKLAFL FAGQGAQVPG MGRGLWEAWP AFRETFDRCV
atepo00a TSPGAVERSIA DSSRGKLAFL FTGQGAQTLG MGRGLYDVWS AFREAFDLCV
atepo04a TPQGAVRGKA VSSRGKLAFL FTGQGAQMPG MGRGLYETWP AFREAFDRCV
atnid07a GIAR..... ..RGRDVAFL FSGQGAQRAG AGRELYASFP VFAQALDEVA
attyl07a GVRD..... ..RDGRMAFL FTGQGSQRAG MAHDLHAAHT FFASALDEVT
atsor02a ~~~~~~AVL FTGQGSQRPT MGRALYDAFP VFRDALDTVA
atsor01a ~~~~~~AIL FTGQGSQRPT MGRALYDAFP VFRGALDAAA
atnys09a GTT.R..... ..AETRLAVL FTGQGAQRLG AGRELAARFP AFATALDAAL
atnys12a GTA..... ..GRGRTAFL FTGQGSQRPG MGRELHDRYP VFADALDEVL
atnys16a GAA.TPH.... ..RT...AFL FSGQGAQRSG MGRELHAAFP VFAAAFDEVV
atnys17a AEA.RER.... ..ST...AFL FSGQGAQRSG MGRELHAAFP VFAAAFDEVV
atnys03a GTA.GEG.... ..PC...AVL FSGQGSQRPG MGRELHARFP VFAAAFDEIT
atnys15a .AA.GRT.... ..RC...AAL FSGQGAQRLG MGRELHARFP VFARALDTAV
atnys07a GRV.GAG.... ..RH...AVL FSGQGAQRLG MGRELYERFP VFAEALDVVV
atnys08a DAV.STG.... ..GS...AVL FTGQGAQRLG MGRELYGRFP VFAEALDVVV
atnys05a GDT.RTG.... ..RH...AVL FSGQGSQRLG MGRELYERFP VFAEALDVAI
atnys06a GTV.TMG.... ..RC...AVL FSGQGSQRLG MGRELYERFP VFAEALDVVI
atnys04a GTV.RTG.... ..RT...AFL FSGQGSQRLG MGRVLYERFP VFAEALDTVL
atnys14a HLQ.GTG.... ..KR...AVL FSGQGSQRLG MGRELHERHP VFAEAFDSVL
atnys00a GAA.HQR.... ..RT...AVL FSGQGSQRPG MGRELAARFP VFADALDDAL
atnys10a EAA.GRG.... ..RT...AVL FTGQGSQRAG MGRELHEVQP EFAAFDAVC
atnys18a GAV.PTG.... ..DRGGLAVL FSGQGSQRPG MGRELHARYP VFAAFDETV
atnys13a AAA.PRT.... ..P.GRTAFL FSGQGAQHAL MGHDLYORFP VYADALDTVL
atavel0a GRT.TSG.... ..ELAVL FAGQGTQRPQ MGRELYEAYP VFAQAIDEIC
atrif02a GVVTTGP.... ..VDGKLAVL FTGQGSQWAG MGRELAETFP VFRDAFEAAC
atmon03a TRAAGPA.... ..RNGGTAFI FTGQGSQRPG MGRQLYDTFD VFAESLDETC
atavel2a GSVGPRGGHS GRRRGKTAML FAGQGTQRPQ MGRQLYAAHP AYADALDQVL
atrif09a GDRAPGVLTG SAKHGKVVYV FPGQGSQRLG MGRELYDRYP VFATAFDEAC
atmon00a AETASIVRGE AYTEGRTAFL FSGQGAQRLG MGRELYAVFP VFADALDEAF
attyl03a PGPRVVTGTA PATERRTAFL FSGQGSQRAG SGRGLYRRHP VFARALDEVC

```

GQG

Fig 2f

10/26

	151		200
atave00x	AAFAPYVDWS	VEQVLRDSPD A.....	PG LDRVDVVQPT
atdebs00p	RAFEPVTDWT	LAQVL.DSPE Q.....	S. .RRVEVVQPA
atepo06p	RAIEAEAGWS	LLGEL.....	.SA.....DEAASQ LGRIDVVQPV
atepo07p	RAIEAEAGWS	LLGEL.....	.SA.....DEAASQ LGRIDVVQPV
atepo01p	RAIQAEAGWS	LLAEL.....	.AA.....DEGSSQ LERIDVVQPV
atepo05p	RAIQAEAGWS	LLAEL.....	.AA.....DETTSQ LGRIDVVQPA
atsoralx	SFIQNLGWS	LLDEL.....	.MT.....DRESSR LDEIDVSLPA
atfkb01p	AVLEPHTGWD	LLDVL.....GEAVV VDRVEVLQPA
atfkb09p	AVLEPYTGWD	LLDVL.....GEAVV AERVEVLQPA
atrap03p	AALSEFVDWD	L.TVL.....DDPAV VDRVDVVQPA
atrap06p	PALREFVDWD	LFTVL.....DDPAV VDRVDVVQPA
atrap04p	AALSEFVDWD	LFAVL.....DDPAV VDRVDVVQPA
atrap13p	AALREFVDWD	LFTVL.....DDPAV VDRVDVVQPA
atrap01p	AALSEFVDWD	LFAVL.....DDPAV VARVDVVQPA
atrap07p	AALSEFVDWD	LFAVL.....DDPAV VDRVDVVQPA
atrap10p	AALSEFVDWD	LFAVL.....DDPAV VDRVDVVQPA
atfkb04x	DALAEHTGRD	LIAML.....DDPAV KSRVDVVHPV
attyl04p	RALSVHVGWD	LLEVVS...GAG LERVDVVQPV
attyl06p	RALSVHVGWD	LLEVVS...GAG LERVDVVQPV
attyl01p	RALSVHVEWD	LLEVVS...GAG LERVDVVQPV
attyl02p	RALSVHVEWD	LLEVVS...GAG LERVDVVQPV
attyl00p	RALSVHVGWD	LLEVVS...GAG LERVDVVQPV
atnid05b	AALRPYLGS	VLSVLRGEPDAPS LDRVDVVQPV
attyl05b	AALSVHLDWS	VLDVLQEKPDAPP LSRVDVVQPV
atnid06x	RALEPHLDWS	VLDVLGAPGTPP IDRADVVPV
atdebs01p	DALEPHLDFE	VIPFLRAEAA RRE.....QDAALS TERVDVVQPV
atmon02p	RALSAHVDWS	LSAVLRG..DGSE LSRVEVVQPV
atmon10p	RALSAYVDWS	LSAVLRG..DGSE LSRVEVVQPV
atmon04p	QALSAYVDWS	LSDVLRG..DGSE LSRVEVVQPV
atmon07p	QALSAYVDWS	LSAVLRG..DGSE LSRVEVVQPV
atmon11p	QALSAYVDWS	LSDVLRG..DGSE LSRVEVVQPV
atmon12p	RALSAHVDWS	LSAVLRG..DGSE LSRVEVVQPV
atmon05b	QALSAYVDWS	LTDVLRG..DGSE LARIDVVQPV
atmon01p	QALSPYVDWS	LTEVLRG..DGRE LSRVDVVQPV
atdebs02p	AVLSEVAGFS	VSEVLEPRPDAPS LERVDVVQPV
atdebs06p	AVLSEVAGFS	ASEVLEQRPDAPS LERVDVVQPV
atave01p	EALAPWVDWS	VVDILRRDAGDAV WERADVVPV
atave07p	EALAPWVDWS	VVDILRRDAGDAV WERADVVPV
atave06p	KALTPWVPWS	LTDILHRDPDDPA WQQADVVPV
atave09p	KALTPWVPWS	LTDILHRDPDDPA WQQADVVPV
atnys01p	AALAEFTDWS	LVDVLRGVVGAPS LERVDVVQPA
atnys11p	AALAEFTDWS	LVDVLRGVVGAPS LERVDVVQPA
atrif05p	AALQPWIDWS	LLDVLRG..EGE. LDRVDVLQPA
atrif07p	AALDQWTDWS	LLDVLRG..DGD. LDSVEVLQPA
atrif08p	AALPWPIDWS	LLDVLRG..EGD. LDRVDVVQPA
atrif10p	AALPWPIDWS	LFDVLRG..DGD. LDRVDVLQPA
atrif03p	AALQWTDWS	LLDVLRG..DADL LDRVDVVQPA
atrif06p	AALPWPIDWS	LLDVLRG..ESDL LDRVDVVQPA
atrif04p	TALGRWVDWS	LTDVLRG..EADL LDRVDVVQPA
atrif01p	AALAPWIDWS	LVDVLRG..EGD. LGRVDVLQPA
atnys02p	TALAPYVDWS	LHDVLRAGEGDPAL LERVDVVQPA
atfkb02p	EALAPYTGWD	LLDVIARRPGAPE LDRVDVVQPA
atave11p	QALEPYLDWS	LLDVLRGAPDAPP LQRVDVVQPV
atdebs03p	ESMAPMQDWK	VSDVLRQAPGAPG LDRVDVVQPV
atnid04p	AALPFFVDWS	LTDVLRGAPGAPG LDRVDVVQPV
atdebs05p	RALAPHVDWS	LTDLL...SGARP LDRVDVVQPA
atdebs04p	EALSPHTDWK	LLDVVRGDGGPDP HERVDVLQPV.

Fig 2g

11/26

atave02a	THLDPHLDHP	LLPLLTQ..N	DNDN.....EDAAAL	LQQTRYAQPA
atave05a	THLDPHLDHP	LLPLLTQNDN	DNDN.....EDAAAL	LQQTPYAQPA
atave04a	THLDPHLDHP	LLPLLTQDPN	TQDT.....	.TTLEEAAL	LQQTPYAQPA
atave08a	THLDPHLDHP	LLPLLTQDPN	TQDT.....	.TTLEEAAL	LQQTPYAQPA
atave03a	THLDPHLDHP	LLPLLTQDPN	TQDT.....	.TTLEEAAL	LQQTRYAQPA
atrap02a	DLLDVP.DLD	VNETGYAQPA
atrap11a	DLLDVP.DLD	VNETGYAQPA
atrap08a	DLLDVP.DLE	VNETGYAQPA
atrap12a	DLLDVP.DLE	VNETGYAQPA
atrap05a	GLLDVP.DLE	VNETGYAQPA
atrap09a	DLLDVP.DLE	VNETGYAQPA
atfkb03a	DALQVPAGLD	VHDTGYAQPA
atfkb07x	GHLN..ADQG	P.....AT
atfkb08x	RRLD...DPD	PHDPTRSQHT
atnid01a	EELQR.....C	GTQNLREVMF	TPD...QPDL	LDRTEYTQPA
atnid03a	EELQR.....C	GTQNLREVMF	TPD...QPDL	LDRTEYTQPA
atnid02a	EELQR.....C	GTQNLREVMF	TPD...QPDL	LDRTEYTQPA
atnid00a	EELQR.....C	GTQNLREVMF	TPD...QPDL	LDRTEYTQPA
atfkb10a	DAFGKHLE..HSPTDVFH	GEHGD....L	AHDTLYAQVG
atrap14a	AGFDKHL..HSLTDVCF	GEGGSTTAGL	VDDTLYAQAG
atmon06a	GLLEGFL...GV	EAGGLREVVF	RGPR....ER	LDHTVWAQAG
atmon08a	GLLEGFL...GV	EAGGLREVVF	RGPR....ER	LDHTMWAQAG
atmon09a	GLLEGFL...GV	GSGGLREVVF	WGPR....ER	LDHTVWAQAG
atepo02a	ALFDRELDRPLREVMW	AEAGSAESLL	LDQTAFTQPA
atepo03x	ALFDRELDRPLREVMW	AEPGSAESLL	LDQTAFTQPA
atepo08a	TLFDRELHQPLCEVMW	AEPGSSRSSL	LDQTAFTQPA
atepo00a	RLFNQELDRPLREVMW	AEPASVDAAL	LDQTAFTQPA
atepo04a	ALFDREIDQPLREVMW	AAPGLAQAAR	LDQTAYAQPA
atnid07a	GGFDAHLERPLLQVMF	AEPGTADAAL	LDRTAYAQPA
atty107a	DRLDPLLGRPLGALLD	ARPGSPEAAL	LDRTEYTQPA
atsor02a	AHLDRDLDRPLRDVLF	APDGSEQAAR	LDQTAFTQPA
atsorbl1a	AHLDRDLDRPLRDVLF	APDGSEQAAR	LDQTAFTQPA
atnys09a	DAFTPHLDRPLREVLWGTDAAL	LDRTAYAQPA
atnys12a	ARLDDGPDRPLREVLW	AAPDSAEEAL	LDRTGYAQPA
atnys16a	AVLDAELGSDAD	GGVSLREVMW	GGG....SEL	LDRTFTQPA
atnys17a	AVLDAELATGSG	GGVSLREVMW	GGG....SEL	LDRTFTQPA
atnys03a	ALLDTHLDRPLREVVW	GTD....ADL	LNDTGWAQPA
atnys15a	DLLDAELGGTLREVIW	GTD....DAP	LNETGFTQPA
atnys07a	DHLDAALPAQAGLREVMW	GDD....AEL	LNETGWTQPA
atnys08a	DHLDAALPAQAGLREVMW	GDD....VEL	LNETGWTQPA
atnys05a	DHLDAALPAQASLREVMW	GDD....VEL	LDETGWTPQA
atnys06a	DHLDAALPAQAGLREVMW	GDD....VEL	LNETGWTQPA
atnys04a	TALDAELGHPLRDIW	GED....AQL	VDRGTGTQPA
atnys14a	ARLDDRLDTPLRDVVW	GTD....EEA	LHATGNTQPA
atnys00a	RALDRHLGDPVREVMW	GTD....AAL	LDRTGWTQPA
atnys10a	AVFDPLLDRPLREVVW	AEDGSDEAAL	LDETGWTPQA
atnys18a	ALLDARL...GTSRLDIVW	DQDRTR....	LDDTRHTQPA
atnys13a	AQFDTVLDPLRAALF	AAPGTPEAAL	LDQTGFTQPA
atave10a	AEADTARTDPGA	PG..LRDVLF	APQDSPEGRL	IEDTGFAQPA
atrif02a	EAVDTHL...	RERPLREVVFDDSAL	LDQTMYTQGA
atmon03a	ARLDPLLEQPLKPVL	APADTAQAAR	LHGTGMTQAA
atave12a	AELDGHLDQPLR	PLIHASADL	..ADVADAADV	LDRTRYAQPA
atrif09a	EQLDVCL..AGR	AGHRVRDVVL	GE.VPAETGL	LNQTVFTQAG
atmon00a	AALDVHLDRP	IREIVLGETD	SGGNVSGENV	IGEGADHQAL	LDQTAYTQPA
atty103a	AALEPHLHRPLRDLMF	AEPGSPEAEP	LDRTEFTQPA

Fig 2h

12/26

	201				250
atave00x	LFAVMISLAA	L.WRSQGVPE	CAVLGHSIGE	IAAAHVSGGL	SLADAARVVT
atdebs00p	LFAVQTSLAA	L.WRSFGVTP	DAVVGHSIGE	LAAAHVCGAA	GAADAARAAA
atepo06p	LFAMEVALSA	L.WRSWGVEP	EAVVGHSMGE	VAAAHVAGAL	SLEDAVAIIC
atepo07p	LFAMEVALSA	L.WRSWGVEP	EAVVGHSMGE	VAAAHVAGAL	SLEDAVAIIC
atepo01p	LFALAVAFAA	L.WRSWGVAP	DVVIGHSMGE	VAAAHVAGAL	SLEDAVAIIC
atepo05p	LFAIEVALSA	L.WRSWGVEP	DAVVGHSMGE	VAAAHVAGAL	SLEDAVAIIC
atsoralx	IISIEIALAA	Q.WRAWGVPE	AFVVGHSSTGE	IAAAHVAGVL	SIEDAMRTIC
atfkb01p	SWAVAVSLAA	L.WQAHGVVP	DAVVGHSQGE	IAAACVAGAL	SLEDAARVVA
atfkb09p	SWAVAVSLAA	L.WQAHGVSP	DAVIGHSQGE	IAAACVAGAL	SLEDAARIVA
atrap03p	SWAVMVSLAA	V.WQAAGVRP	DAVIGHSQGE	IAAACVAGAV	SLRDAARIVT
atrap06p	SWRMMVSLAA	V.WQAAGVRP	DAVIGHSQGE	IAAACVAGAV	SMRDAARIVT
atrap04p	SWAVMVSLAA	V.WQAAGVRP	DAVIGHSQGE	IAAACVAGAV	SLRDAARIVT
atrap13p	SWAMMVSLAA	V.WQAAGVRP	DAVIGHSQGE	IAAACVAGAV	SLRDAARIVT
atrap01p	SWAVMVSLAA	V.WQAAGVRP	DAVVGHSQGE	IAAACVAGAV	SLRDAARVVT
atrap07p	SWAVMVSLAA	V.WQADGVVP	DAVIGHSQGE	IAAACVAGAV	SLRDAARSVT
atrap10p	SWAVMVSLAA	V.WQAAGVRP	DAVIGHSQGE	IAAACVAGAV	SMRDAARIVT
atfkb04x	CWAVMVSLAA	V.WEAAGVRP	DAVVGHSQGE	IAAACVAGAI	SLEDGARLVA
attyl04p	TWAVMVSLAR	Y.WQAMGVDV	AAVVGHSQGE	IAAATVAGAL	SLEDAAAVVA
attyl06p	TWAVMVSLAR	Y.WQAMGVDV	AAVVGHSQGE	IAAATVAGAL	SLEDAAAVVA
attyl01p	TWAVMVSLAR	Y.WQAMGVDV	AAVVGHSQGE	IAAATVAGAL	SLEDAAAVVA
attyl02p	TWAVMVSLAR	Y.WQAMGVDV	AAVVGHSQGE	IAAATVAGAL	SLEDAAAVVA
attyl00p	TWAVMVSLAR	Y.WQAMGVDV	AAVVGHSQGE	IAAATVAGAL	SLEDAAAVVA
atnid05b	LFTMMVSLAA	V.WRALGVPE	AAVVGHSQGE	IAAAHVAGAL	SLDDSAARIVA
attyl05b	LFTMMLSLAA	C.WRDLGVHP	AAVVGHSQGE	IAAACVAGAL	SLEDAARIVA
atnid06x	LFTTMVSLAA	L.WEAHGVVP	AAVVGHSQGE	VAAACVAGAL	SLDDAALVIA
atdebs01p	MFAVMVSLAS	M.WRAHGVPE	AAVIGHSQGE	IAAACVAGAL	SLDDAARVVA
atmon02p	LWAVMVSLAA	V.WADYGVTP	AAVIGHSQGE	MAAACVAGAL	SLEDAARIVA
atmon10p	LWAVMVSLAA	V.WADYGVTP	AAVIGHSQGE	MAAACVAGAL	SLEDAARIVA
atmon04p	LWAVMVSLAA	V.WADYGVTP	AAVVGHSQGE	MAAACVAGAL	SLEDAARIVA
atmon07p	LWAVMVSLAA	V.WADYGVTP	AAVIGHSQGE	MAAACVAGAL	SLEDAARVVA
atmon11p	LWAVMVSLAA	V.WADYGVTP	AAVIGHSQGE	MAAACVAGAL	SLEDAARIVA
atmon12p	LWAVMVSLAS	V.WADYGVTP	AAVIGHSQGE	MAAACVAGAL	SLEDAARIVA
atmon05b	LWAVMVALAA	V.WADQGIEP	AAVVGHSQGE	IAAACVVGAI	SLDEAARIVA
atmon01p	LWAVMVSLAA	V.WADHGVTP	AAVVGHSQGE	IAAVVVGAL	TLEDGAKIVA
atdebs02p	LFAVMVSLAR	L.WRACGAVP	SAVIGHSQGE	IAAAVVAGAL	SLEDGMRVVA
atdebs06p	LFSVMVSLAR	L.WGACGVSP	SAVIGHSQGE	IAAAVVAGVL	SLEDGVRVVA
atave01p	LFSVMVSLAA	L.WRSYGIEP	DAVLGHSQGE	IAAAHVCGAL	SLKDAAKTVA
atave07p	LFSVMVSLAA	L.WRSYGIEP	DAVLGHSQGE	IAAAHVCGAL	SLKDAAKTVA
atave06p	LFSIMVSLAA	L.WRSYGIEP	DAVLGHSQGE	IAAAHICGAL	SLKDAAKTVA
atave09p	LFSIMVSLAA	L.WRSYGIEP	DAVLGHSQGE	IAAAHICGAL	SLKDAAKTVA
atnys01p	SFAVMVSLAA	L.WGSRGVLP	DAVVGHSQGE	IAAAVVSGAL	SLRDGARVVA
atnys11p	SFAVMVSLAA	L.WRSRGVLP	DAVVGHSQGE	IAAAVVSGAL	SLRDGARVVA
atrif05p	CFAVMVGLAA	V.WASVGVVP	DAVLGHSQGE	IAAACVSGAL	SLEDAAKVVA
atrif07p	CFAVMVGLAA	V.WESAGVRP	DAVVGHSQGE	IAAACVSGAL	TLDDAAKVVA
atrif08p	SFAVMVGLAA	V.WSSVGVVP	DAVLGHSQGE	IAAACVSGAL	SLQDAAKVVA
atrif10p	CFAVMVGLAA	V.WSSAGVVP	DAVLGHSQGE	IAAACVSGAL	SLEDAAKVVA
atrif03p	SFAMMVGLAA	V.WTSLGVTP	DAVLGHSQGE	IAAACVSGAL	SLDDAAKVVA
atrif06p	SFAMMVGLAA	V.WQSVGVVP	DAVVGHSQGE	IAAACVSGAL	SLQDAAKVVA
atrif04p	SFAVMVGLAA	V.WASLGVPE	EAVVGHSQGE	IAAACVSGAL	SLEDAAKVVA
atrif01p	CFAVMVGLAA	V.WESVGVVP	DAVVGHSQGE	IAAACVSGAL	SLEDAAKVVA
atnys02p	LFAMMVGLSA	L.WRSHGVVP	AAVVGHSQGE	IAAACVAGAL	SLADAARVVA
atfkb02p	SFAMMVGLAE	L.WRAHGVAP	AAVVGHSQGE	VAAACVAGVL	TLDDAAKVVA
atave11p	LFAVMVSLAA	L.WRSYGVHP	DAVAGHSQGE	IAAAHVAGAL	SLDDAARVTA
atdebs03p	LFAVMVSLAE	L.WRSYGVPE	AAVVGHSQGE	IAAAHVAGAL	TLEDAAKLVV
atnid04p	TFAVVVALAA	M.WRWLGVEP	AAVVGHSQGE	IAAAHVAGVL	SLEDAARVVA
atdebs05p	LFAVMVSLAA	L.WRSHGVPE	AAVVGHSQGE	IAAAHVAGAL	TLEDAAKLVA
atdebs04p	LFSIMVSLAE	L.WRAHGVTP	AAVVGHSQGE	IAAAHVAGAL	SLEAAAKVVA

Fig 2i

13/26

```

atave02a QRATLMQTMP P..GTMTTLH TTPHHIT..H HLTAE...N DLAIAAINTP
atave05a QRATLMQTMP P..GTMTTLH TTPHHIT..H HLTAE...N DLAIAAINTP
atave04a QRATLMQTMP P..GTMTTLH TTPHHIT..H HLTAE...N DLAIAAINTP
atave08a QRATLMQTMP P..GTMTTLH TTPHHIT..H HLTAE...N DLAIAAINTP
atave03a QRATLMQTMP P..GTMTTLH TTPHHIT..H HLTAE...N DLAIAAINTP
atrap02a ARARLMQALP AG.GVMAAVP VSEDEARAVL G.....E GVEIAAVNGP
atrap11a ARARLMQALP AG.GVMVAVP VSEDEARAVL G.....E GVEIAAVNGP
atrap08a ARARLMQALP AG.GVMVAVP VSEDEARAVL G.....E GVEIAAVNGP
atrap12a ARARLMQALP AG.GVMVAVP VSEDEARAVL G.....E GVEIAAVNGP
atrap05a ARARLMQALP PG.GVMVAVP VSEDEARAVL G.....E GVEIAAVNGP
atrap09a ARARLMQALP AG.GVMVAVP VSEDEARAVL G.....E GVEIAAVNGP
atfkb03a ARARLMQALP PG.GAMAAVS ASERDALPLL C.....E GVEIAAVNGP
atfkb07x ARSRIMDELP TG.GAMVTVL TSEENALRAL R.....P GVEIAAVNGP
atfkb08x TRARLMHTLP PP.GAMVTVL TSEEARQAL R.....P GVEIAAVFGP
atnid01a ARAHLMGQLP HG.GAMLSVQ AAEHDLDQLA ....HT...H GVEIAAVNGP
atnid03a ARAHLMGQLP HG.GAMLSVQ AAEHDLDQLA ....HT...H GVEIAAVNGP
atnid02a ARAHLMGQLP HD.GAMLSVQ AAEHDLDQLA ....HT...H GVEIAAVNGP
atnid00a ARAHVMGQLP HG.GAMLSVQ AAEHDLDQLA ....HT...H GVEIAAVNGP
atfkb10a ARGRALRALP P..GAMTAVE GSPAEEVG..A FTD..... .LDIAAVNGP
atrap14a ARGRALRTTP P..GAMVALR AGEEEVR..E FLSRTG...A ALDLAAVNSP
atmon06a ARARLMGGLP EG.GAMCAVQ ATPAELAA.. .DVDG...S AVSVAAVNTP
atmon08a ARARLMGGLP EG.GAMCAVQ ATPAELAA.. .DVDD...S GVSVAAVNTP
atmon09a ARARLMGGLP EG.GAMCAVQ ATPAELAA.. .DVDG...S SVSVAAVNTP
atepo02a ARGRLMQGLS AG.GAMVSLG APEAEVA..A AVAPHA...A SVSIAAVNGP
atepo03x ARGRLMQGLS AG.GAMVSLG APEAEVA..A AVAPHA...A SVSIAAVNGP
atepo08a ARGRLMQALP AG.GAMVSIA APEADVA..A AVAPHA...A LVSIAAVNGP
atepo00a ARGRLMQALP AG.GAMVSIE APEADVA..A AVAPHA...A SVSIAAVNAP
atepo04a ARGRLMQALP AG.GAMVAIA ASEAEVA..A SVAPHA...A TVSIAAVNGP
atnid07a ARGRLMQRLP EG.GAMVAVR ATEQEVAELE WIAGGR.... AV.VAAFNGP
atlyl07a ARGRLMQRLP PG.GAMVSVR AGEDEVRL.. .LAGRE...D AVCVAAVNGP
atsor02a ARAKLMQALP QG.GAMVTLR ASEEEVRDL.. .LQPYD...G RASLAALNGP
atsorbl1a ARAKLMQALP QG.GAMVTLQ ASEQEARDL.. .LQAAE...G RVSLAAVNGH
atnys09a ARGRLMQALP DG.GAMIAVQ ASEADVAPL.. .LAGHE...D QVAIAAVNGP
atnys12a ARGRLMQALP EG.GAMVALE AAEDEVLP.. .LEGLT...D RVSVAAVNGP
atnys16a ARASLMDALP VG.GVMVAVE AAEAEVVP.. .L...V...D GVAIAAVNGP
atnys17a ARASLMDALP VG.GVMVAVE AAEAEVVP.. .L...V...D GVAIAAVNGP
atnys03a ARARLMQALP RG.GAMLAIR ATEDEVTPH.. .L...T...D DVSIAAVNGP
atnys15a ARAGLMQALP RG.GAMVAVE ATEDEVSP.. .L...T...D GVAIAAINGP
atnys07a ARATLMQALP AG.GAMIAVQ ATEDEVTPH.. .L...T...D DVAIAAINGP
atnys08a ARATLMQALP TG.GAMIAVQ ATEDEVTPH.. .L...T...D EVAIAAVNGP
atnys05a ARATLMQALP TG.GAMIAIQ AAEDEVTOH.. .L...T...D DVSIAAVNGP
atnys06a ARATLMQALP AG.GAMIAVQ ATEDEVIPH.. .L...T...D EVAIAAVNGP
atnys04a ARAVLMQSLP EG.GAMIAVQ ATEDEVLP.. .L...T...D DVSIAAVNSP
atnys14a ARAALMQRLP AG.GAMIAVE ATEDEVTP.. .L...T...D GVSIAAVNGP
atnys00a ARATLMQALP AG.GAMAALE ATEDEVAPL.. .L...G...A HLALAAVNGP
atnys10a ARATLMQALP TG.GAMIAIQ ATEDEIAAH.. .L...D...D TVAIAAVNGP
atnys18a ARATAMSELP PG.GAMVALE ATEDEVRL.. .L...T...D DLAAIAAVNAP
atnys13a ARASLMQQLP RD.GAMVALE ATEDEVAPL.. .L...T...D GVALAAVNGP
atave10a ARGRLMQGLP SG.GAMVAIE ASEDEILPL.. .PDEYA...S RVAAHAAVNGP
atrif02a ARGRLMQALP AG.GAMVAVQ ATEDEVAPL.. .LDGT..... .VCVAAVNGP
atmon03a ARGRLMQALP AG.GAMLAVQ AAEEDVLPL.. .LAGQE...E RLSLAAVNGP
atave12a ARGRLMQQLA PG.GAMVAVR ASEAEAR..Q ALDGRE...A RVSVAAVNGP
atrif09a ARGRLMQALA PG.GAMVAVA ASEAEVAELL G.....D GVELAAVNGP
atmon00a TRGRLMQAVR AP.GAMAAWQ ATADEAA..E QLAGHE...R HVTVAAVNGP
atlyl03a ARGRLMQALP AG.GAMAAIR ATAEEIAPL.. .LERRA...G ELALAAVNGP

```

*

Arginine

Fig 21

14/26

	301		350
atave00x	RSTVVSAGARE	AVADLVADLT	AAQVTRMIP .VDVPAHSPL MYAIEERVV. Load AT
atdebs00p	RSVLLTGSPE	PVARRVQELS	AEGVRAQVIN .VSMAAHSQA VDDIAEGMR. Load AT
atepo06p	RSTVLAGEPA	ALSEVLAAIT	AKGVFWRQV. KVDVASHSPQ VDPLREEL.I
atepo07p	RSTVLAGEPA	ALSEVLAAIT	AKGVFWRQV. KVDVASHSPQ VDPLREEL.I
atepo01p	RSTVLSGEP	AIGEVLSSLN	AKGVFCRRV. KVDVASHSPQ VDPLREEL.I
atepo05p	RSTVLAGEPA	ALAEVLAILA	AKGVFCRRV. KVDVASHSPQ IDPLRDEL.L
atsoralx	DSTVLAGEPD	ALDALLQALE	RKNVFCRRV. AMDVAPHCPQ VDCLRDEL.F Benzoate-CoA
atfkb01p	ESTVVAGDPA	AVERVLARYE	AEGVRVRRV. AVDYASHTPH VEAIEAQL.A
atfkb09p	ESTVVAGDPS	AVERVLARYE	AEGVRVRRV. AVDYASHTPH VEAIEAQL.A
atrap03p	ASTVIAGTPE	AVDHVLTAE	ARGVRVRRV. TVDYASHTPH VELIRDEL.L
atrap06p	ASTVIAGTPE	AVDHVLTAE	ARGVRVRRV. TVDYASHTPH VELIRDEL.L
atrap04p	ASTVIAGTPE	AVDHVLTAE	ARGVRVRRV. TVDYASHTPH VELIRDEL.L
atrap13p	ASTVIAGTPE	AVDHVLTAE	AQGVRRVRRV. TVDYASHTPH VELIRDEL.L
atrap01p	ASTVVAGAPE	AVDRVLAVHE	ARGVRVRRV. AVDYASHTPH VELIRDEL.L
atrap07p	ASTVVAGAPE	AVDRVLAVHE	ARGVRVRRV. AVDYASHTPH VELIRDEL.L
atrap10p	ASTVIAGTPE	AVDHVLTALR	QAGAGAAD. . HVDYASHTPH VELIRDEL.L
atfkb04x	ATTIVSGRPD	AVETLIADYE	ARGVWVTRL. VVDCPTHPTF VDFLYDEL.Q C5 unit
attyl04p	ASTVVSDDR	AVAGYVAVCQ	AEGVQARLIP .VDYASHSRH VEDLKGELE.
attyl06p	ASTVVSDDR	AVAGYVAVCQ	AEGVQARLIP .VDYASHSRH VEDLKGELE.
attyl01p	ASTVVSDDR	AVAGYVAVCQ	AEGVQARLIP .VDYASHSRH VEDLKGELE.
attyl02p	ASTVVSDDR	AVAGYVAVCQ	AEGVQARLIP .VDYASHSRH VEDLKGELE.
attyl00p	ASTVVSDDR	AVAGYVAVCQ	AEGVQARLIP .VDYASHSRH VEDLKGELE.
atnid05b	GSCAVAGDPE	ALAEVLALLT	GEGVHARPIP GVDTAGHSPQ VDALRAHL.L Etmalonyl-CoA
attyl05b	GTAAGVAGD	ALRELLAELT	AEGIRAKPIP GVDTAGHSAQ VDLKEHL.F Etmalonyl-CoA
atnid06x	ASVTVSGDAL	ALAEFGARLS	AEGVLRWPLP GVDFAHSPQ VEEFRAEL.L MeOmalonyl-CoA
atdebs01p	RSVVVAGDS	ELDRVLASCT	TECIRAKRL. AVDYASHSSH VETIRDALHA
atmon02p	SSTVISGPPE	HVAAVVADAE	ARGLRARVID .VGYASHGPQ IDQLHDL.L.T
atmon10p	SSTVISGPPE	HVAAVVADAE	ARGLRARVID .VGYASHGPQ IDQLHDL.L.T
atmon04p	SSTVISGPPE	HVAAVVADAE	ARGLRARVID .VGYASHGPQ IDQLHDL.L.T
atmon07p	SSTVISGPPE	HVAAVVADAE	ERGLRARVID .VGYASHGPQ IDQLHDL.L.T
atmon11p	SSTVISGPPE	HVAAVVADAE	AQGLRARVID .VRYASHGPQ IDQLHDL.L.T
atmon12p	SSTVISGPPE	HVAAVVADAE	ARGLRARVID .VGYASHGPQ IDQLHDL.L.T
atmon05b	SSTVISGPPE	GIAAVVADAQ	ERGLRARAVA .SDVAGHGPQ LDAILDQL.T Et/mal-CoA
atmon01p	SSTVISGPPE	QVAAVVADAE	ARELRGRVID .VDYASHSPQ VDAITDEL.T
atdebs02p	DAVVVAGDAQ	AAREFLEYCE	GVGIRARAIP .VDYASHTAH VEPVRDEL.V
atdebs06p	SSVVVSGDPE	ALAEVLARCE	DEGVRAKTLF .VDYASHSRH VEEIRETI.L
atave01p	RSTAVSGDAE	AVDEVLAYCA	GTGVRARRIP .VDYASHCPH VQPLREEL.L
atave07p	RSTAVSGDAE	AVDEVLAYCA	GTGVRARRIP .VDYASHCPH VQPLREEL.L
atave06p	HSTTVSGDTK	AVDEVLAHCT	DTGLRAKRIP .VDYASHCPH VQPLHDEL.L
atave09p	HSTTVSGDTT	AVEELLTHCA	DTGLRAKRIP .VDYASHCPH VQPLHDEL.L
atnys01p	RSVVVAGEPE	ALDALHARLT	ADDIRARRIA .VDYASHSHQ VEDLHEEL.L
atnys11p	RSVVVAGEPE	ALDALHARLT	ADDIRARRIA .VDYASHSHQ VEDLHEEL.L
atrif05p	ASVVIAGDAE	ALTEAVEVLG	G.RRVA .VDYASHTRH VEDIQDTL.A
atrif07p	ASVVIAGDAQ	ALDEALEVLA	GDGVRVRQVA .VDYASHTRH VEDIRDTL.A
atrif08p	SSVVIAGDAE	ALDQALEALT	GQDIRVRRVA .VDYASHTRH VEDIQDTL.A
atrif10p	ASVVIAGDAQ	ALDETLEALS	GAGIRARRVA .VDYASHTRH VEDIEDTL.A
atrif03p	SSVVIAGDAQ	ALDEALEALA	GDGVRVRRVA .VDYASHTRH VEAIAETL.A
atrif06p	ASVVIAGEAQ	ALDEVVDALS	GQEVRRVRRVA .VDYGSHTNQ VEAIEDLL.A
atrif04p	TSVVIAGDAE	ALDEALDALD	DQGVRIARRVA .VDYASHTRH VEAARDAL.A
atrif01p	SSVVIAGDAH	ALDATLEILS	GEGIRVRRVA .VDYASHTRH VEDIRDTL.A
atnys02p	SSVVVSGDTD	ALDALHTACQ	EQGVRRARKVS .VDYASHGRH VEAVRDEL.A
atfkb02p	ASIVVAGAAD	AVEELLAATPHARRIA .VDYASHTAH VESIRGAL.L
atave11p	RSVVVSGEPE	AVDALVEELS	HEDVPARRLM .VDWASHSPQ VEAIQGRL.L
atdebs03p	RSVVVSGEPG	ALRAFSEDCA	AEGIRVRDID .VDYASHSPQ IERVREEL.L
atnid04p	ETTVVCGAPG	AVDSLGLVLQ	GEGVRVRRID .VDYASHSRH VEGIRDEL.A
atdebs05p	RSVVVAGESG	PLDELIAECE	AEGITARRIP .VDYASHSPQ VESLREEL.L
atdebs04p	GTSVVAGPTA	ELDEFFAEAE	AREMKPRRIA .VRYASHSPE VARIEDRL.A

Fig 2m

15/26

atave02a	TSLVISGTPH	TVQHITTLCQ	QQGIKTKTL.	PTNHAFHSPH	TNPILNQLH.	
atave05a	TSLVISGTPH	TVQHITTLCQ	QQGIKTKTL.	PTNHAFHSPH	TNPILNQLH.	
atave04a	TSLVISGTPH	TVQHITTLCQ	QQGIKTKTL.	PTKNAFHSPH	TNPILNQLH.	
atave08a	TSLVISGTPH	TVQHITTLCQ	QQGIKTKTL.	PTNHAFHSPH	TNPILNQLH.	
atave03a	TSLVISGTPH	TVQHITTLCQ	QQGIKTKTL.	PTNHAFHSPH	TNPILNQLH.	
atrap02a	SSVVLSGDEA	AVLQAAEGLGKWTRL.	PTSHAFHSAR	MEPMLLEEFR.	
atrap11a	SSVVLSGDEA	AVLQAAEGLGKWTRL.	ATSHAFHSAR	MEPMLLEEFR.	
atrap08a	SSVVLSGDEA	AVLQAAEGLGKWTRL.	ATSHAFHSAR	MEPMLLEEFR.	
atrap12a	SSVVLSGDEA	AVLQAAEGLGKWTRL.	ATSHAFHSAR	MEPMLLEEFR.	
atrap05a	SSVVLSGDET	AVLQAAAALGKSTRL.	ATSHAFHSAR	MEPMLLEEFR.	
atrap09a	SSVVLSGDEA	AVLQAAEGLGKWTRL.	ATSHAFHSAR	MEPMLLEEFR.	
atfkb03a	ASIVLSGDED	AVLDVAARLGRFTRL.	RTSHAFHSAR	MEPMLDEFR.	
atfkb07x	HSVVLSGDEG	PVLDVAQQLGIHHRL.	PTRHAGHSAR	MDPLVAPLL.	MeOmalonyl-CoA
atfkb08x	HSVVLSGDED	AVLDVAQRLGIHHRL.	PAPHAGHSAH	MEPVAAELL.	MeOmalonyl-CoA
atnid01a	THCVLSGPRT	ALEETAQQLH	QQGIRHTWL.	KVSHAFHSAL	MDPMLGAFR.	
atnid03a	THCVLSGPRT	ALEETAQHLR	EQNVRHTWL.	KVSHAFHSAL	MDPMLGAFR.	
atnid02a	THCVLSGPRT	ALEETAQHLR	EQNVRHTWL.	KVSHAFHSAL	MDPMLGAFR.	
atnid00a	THCVLSGPRT	ALEETAQHLR	EQNVRHTWL.	KVSHAFHSAL	MDPMLGAFR.	
atfkb10a	SAVVLTGAPD	DVAAFERAWA	AAGRRAKRL.	DVGHAHFSRH	VDGALDDFR.	
atrap14a	EAVVVSGEPE	PVADFEAAWT	ASGREARKL.	KVRHAFHSRH	VEAVLDEFR.	
atmon06a	DSTVISGPSD	EVDRIAGVWR	ERGRKTKAL.	SVSHAFHSAL	MEPMLAEFT.	
atmon08a	DSTVISGPSG	EVDRIAGVWR	ERGRKTKAL.	SVSHAFHSAL	MEPMLAEFT.	
atmon09a	DSTVISGPSG	EVDRIAGVWR	ERGRKTKAL.	SVSHAFHSAL	MEPMLGEFT.	
atepo02a	EQVVIAGVEQ	AVQAIAGFA	ARGARTKRL.	HVSHAFHSPL	MEPMLLEFG.	
atepo03x	EQVVIAGVEQ	AVQAIAGFA	ARGARTKRL.	HVSHAFHSPL	MEPMLLEFG.	Mal/mmal
atepo08a	EQVVIAGAEK	FVQQIAAFA	ARGARTKPL.	HVSHAFHSPL	MDPMLAEFR.	
atepo00a	DQVVIAGAGQ	PVHAIAAAMA	ARGARTKAL.	HVSHAFHSPL	MAPMLAEFG.	
atepo04a	DAVVIAGAEV	QVLALGATFA	ARGIRTKRL.	AVSHAFHSPL	MDPMLDEFQ.	
atnid07a	DSLVLSGDEQ	AVVSAAGELA	ARGRRTKRL.	SVSHAFHSPH	MDAMLADFR.	
attyl07a	RSVVISGAEE	AVAEAAAQLA	GRGRRTTRL.	RVSHAFHSPL	MDGMLAGFR.	
atsor02a	LSTVVAGDED	AVVEIARQAE	ALGRKTTTRL.	RVSHAFHSPH	MDGMLDDFR.	
atsorbl1a	LSTVVAGDED	AVLKIARQVE	ALGRKATRL.	RVSHAFHSPH	MDGMLDDFR.	
atnys09a	SAVVLSGAEE	TVTALAEQLA	ADGRKTRRL.	RVSHAFHSPL	MEPMLDAFR.	
atnys12a	RSVVVAGVEE	DVLLLDLFA	ADGRRTKRL.	RVSHAFHSPL	MDAMLDDFA.	
atnys16a	VSVVVSVEA	AVGQVVDQLV	ERGRRVRL.	AVSHAFHSPL	MDPMLDAFR.	
atnys17a	VSVVVSVEA	AVGQVVDQLV	ERGRRVRL.	AVSHAFHSPL	MDPMLDAFR.	
atnys03a	TSVVVAGTEE	AVAAIGARFT	AQDRKTTTRL.	RVSHAFHSPL	MDPMLAEFR.	
atnys15a	TSLVVSGET	ATLAVARLA	EQGRRTTRL.	RVSHAFHSPL	MDPMLAEFR.	
atnys07a	NALVVSGVED	AAVEIGARFA	AEGRRTRRL.	HVSHAFHSPL	MDPMLAEFR.	
atnys08a	TSVVISGAEE	ATQTVAQHFA	DQGRRTTAL.	RVSHAFHSPL	MDPMLAEFR.	
atnys05a	TSVVVSGAES	AARTVADRLA	ENGRKTTTRL.	RVSHAFHSPL	MDPMLAEFR.	
atnys06a	TSVVISGAEE	ATQTVAQHFA	DQGRRTTAL.	RVSHAFHSPL	M. .MLAEFR.	
atnys04a	TSVVVSGYEN	ATLAVARHFA	DQGRRTTRL.	RVSHAFHSPL	MAPMLDDFR.	
atnys14a	TAVVLSGAGD	AVTALGQALA	ERGHRTTRL.	RVSHAFHSHL	MDPMLADFR.	
atnys00a	TAVVVAGAED	AVRQLTARFA	DRGRRTSRL.	AVSHAFHSPL	MEPMLDAFR.	
atnys10a	QSVVISGDEE	AAETIAATFA	ERGRKTKRL.	RVSHAFHSPL	MDGMLDAFR.	
atnys18a	RSVVVAGAED	AALAVRRHFD	DLGRRTTRL.	PVSHAFHSPL	MDPMLDAFR.	
atnys13a	RSVVVAGAED	AVRAVADRLA	ADGRRTTRL.	TVSHAFHSPL	MDPMLTDF.	
atave10a	RSIVLSGDED	AVLDLAQQWA	ARGRRTRRL.	RTSHAFHSPH	MDAMLGDFR.	
atrif02a	DSVVLSGTEA	AVLAVADELA	GRGRKTRRL.	AVSHAFHSPL	MEPMLDDFR.	
atmon03a	TAVVVSGEAA	AVGEVEKALR	GRGLKTKRL.	NVSHAFHSPL	IEPMLDDFR.	
atave12a	ASVVFSGAED	EVGNMADWFA	ERGRRVKRL.	RTGHAFHSPL	MDPMLLEFQ.	
atrif09a	SAVVLSGDAD	AVVAAAARMR	ERGHKTKQL.	KVSHAFHSAR	MAPMLAEFA.	
atmon00a	DSVVVSGDRA	TVDELTAAWR	GRGRKAHHL.	KVSHAFHSPH	MDPILDEL.	
attyl03a	SSVVVSGDEA	AVLELLEQWR	AEGREARRL.	AVSHAFHSPL	MDGMLTQFD.	

**** HAFH/YASH/TAGH motif

Fig 2n

16/26

	351		400
atave00x	SGLLPITPRP	SRIPFHSSVT	G....GRL. .DTRELDAA YWYRNMSSTVR
atdebs00p	SALAWFAPGG	SEVPFYASLT	G....GAV. .DTRELVADY WRRSFLRPVR
atepo06p	AALGAIRPRA	AAVPMRSTVT	G....GVI. .AGPELGASY WADNLRQPVVR
atepo07p	AALGAIRPRA	AAVPMRSTVT	G....GVI. .AGPELGASY WADNLRQPVVR
atepo01p	AALGGLRPGA	AAVPMRSTVT	G....AMV. .AGPELGANY WMNNLRQPVVR
atepo05p	AALGELEPRQ	ATVSMRSTVT	S....TIM. .AGPELVASY WADNVRQPVVR
atsoralx	DALREVRPNK	AQIPIVSEVT	G....TAL. .DGERFDASH WVRNFGDPAL
atfkb01p	DALEGITSSST	PSVPWWSTVD	S....GWV. .TEPFGDAY WYRNLQRPVA
atfkb09p	DVLGDITSSA	PSVPWWSTVD	G....GWV. .TEPAGDDY WYRNLQRPVA
atrap03p	DITSDDSSSQ	PLVPWLSTVD	G....SWV. .DSPLDGEY WYRNLREPVG
atrap06p	DITSDDSSSQ	PVVPWLSTVD	G....SWV. .DSPLDVEY WYRNLREPVG
atrap04p	GITAGIGSQP	PVVPWLSTVD	G....SWV. .DSPLDGEY WYRNLREPVG
atrap13p	DITSDDSSQT	PLVPWLSTVD	G....TWV. .DSPLDGEY WYRNLREPVG
atrap01p	GVIAGVDSRA	PVVPWLSTVD	G....TWV. .EGPLDAEY WYRNLREPVG
atrap07p	DITAGIGSQ	PVVPWLSTVD	G....TWV. .EGPLDVEY WYRNLREPVG
atrap10p	DITSDDSSQD	PLVPWLSTVD	G....TWV. .DSPLDGEY WYRNLREPVG
atfkb04x	RIVAATTSRA	PEIPWFSTAD	E....RWI. .DAPLDDEY WFRNMRNPVG
atlyl04p	RVLSGIRPRS	PRVPVCSTVA	G....E..Q PGEPVFDAGY WFRNLNRNVE
atlyl06p	RVLSGIRPRS	PRVPVCSTVA	G....E..Q PGEPVFDAGY WFRNLNRNVE
atlyl01p	RVLSGIRPRS	PRVPVCSTVA	G....E..Q PGEPVFDAGY WFRNLNRNVE
atlyl02p	RVLSGIRPRS	PRVPVCSTVA	G....E..Q PGEPVFDAGY WFRNLNRNVE
atlyl00p	RVLSGIRPRS	PRVPVCSTVA	G....E..Q PGEPVFDAGY WFRNLNRNVE
atnid05b	EVLAPVAPRP	ADIPFYSTVT	G....GLL. .DGTELDATY WYRNMREPVE
atlyl05b	EVLAPVSPRS	SDIPFYSTVT	G....APL. .DTERLDAGY WYRNMREPVE
atnid06x	DLISGVRPAP	SRIPFFSTVT	A....GPC. .GGDQLDGAY WYRNTREPVE
atdebs01p	ELGEDFHPLP	GFVFFFSTVT	G....RWT. .QPDELDAGY WYRNLRRTVR
atmon02p	ERLADIRPTN	TDVAFYSTVT	A....ERL. .TDTTALDIDY WVTNLQRPVR
atmon10p	ERLADIRPAN	TDVAFYSTVT	A....ERL. .TDTTALDIDY WVTNLQRPVR
atmon04p	EGLADIRPAN	TDVAFYSTVT	A....ERL. .TDTTALDIDY WVTNLQRPVR
atmon07p	DRLADIRPAT	TDVAFYSTVT	A....ERL. .TDTTALDIDY WVTNLQRPVR
atmon11p	DRLADIQPTT	TDVAFYSTVT	A....ERL. .DDTTALDITAY WVTNLQRPVR
atmon12p	ERLADIRPTT	TDVAFYSTVT	A....ERL. .DDTTALDIDY WVTNLQRPVR
atmon05b	EGLAGIRPAA	TDVAFYSTVT	A....GHL. .TDTTELDITAY WVRNVRRTVR
atmon01p	HTLSGVRPTT	APVAFYSAVT	G....TRI. .DTAGLDIDY WVTNLRRPVR
atdebs02p	QALAGITPRR	AEVFFFSTLT	G....D..F LDGTELDAGY WYRNLRRPVE
atdebs06p	ADLDGISARR	AAIPLYSTLH	G....E..R RD...MGPRY WYDNLRSQVR
atave01p	ELLGDISPQP	SGVFFFSTVE	G.....TW LDTTTLDAAY WYRNLHQPVVR
atave07p	ELLGDISPQP	SGVFFFSTVE	G.....TW LDTTTLDAAY WYRNLHQPVVR
atave06p	HLLGDITPQP	STVFFFSTVE	G.....TW LDTTTLDAAY WYRNLHQPVVR
atave09p	HLLGDITPQP	STMPFFSTVV	G....HLVW Y.TTTTLDAAY WYRNLHQPVVR
atnys01p	EVLAEELAPRT	SEVFFFSTVT	G....DWL. .DTARMDAGY WFRNLRRGRVR
atnys11p	EVLAEELAPRT	SEVFFFSTVT	G....DWL. .DTARMDAGY WFRNLRRGRVR
atrif05p	ETLAGIDAQA	PVVPFYSTVA	G....EWI. .TDAGVVDGGY WYRNLRRQVG
atrif07p	ETLAGITAQA	PDVFFRSTVT	G....GWV. .RDADVLDDGY WYRNLRRQVR
atrif08p	EALAGIEAHA	PTLPFFSTLT	G....DWI. .REAGVVDGGY WYRNLRRQVG
atrif10p	EALAGIDARA	PLVPFLSTLT	G....EWI. .RDEGVVDGGY WYRNLRRGRVR
atrif03p	KTLAGIDARV	PAIPFYSTVL	G....TWI. .EQA.VVDAGY WYRNLRRQVR
atrif06p	ETLAGIEAQA	PKVPPFYSTLI	G....DWI. .RDAGIVDDGY WYRNLRRQVG
atrif04p	EMLGIGIRAQA	PEVPPFYSTVT	G....GWV. .EDAGVLDDGY WYRNLRRQVR
atrif01p	ETLAGISAQA	PAVPPFYSTVT	S....EWV. .RDAGVLDDGY WYRNLRRQVR
atnys02p	RVLAPVDPRA	PEVPPFYSTVT	G....DRV. .DDAA.FDGAY WYTNLRQTVR
atfkb02p	DALADLTPGA	PEIPFFSTVD	E....AWL. .DRPA..DAAY WYDNVRCFVR
atavellp	ELLAPIRART	GDVPPFYSTVT	G....ERI. .DGTELDADY WYRNLRRQVR
atdebs03p	ETTGDIAPRP	ARVTFHSTVE	S....RSM. .DGTELDARY WYRNLRETVR
atnid04p	AVLAGLRPRA	GRVPPFYSTVE	A....EPL. .DGTALDAGY WYRNLRRQVR
atdebs05p	TELAGISPV	ADVALYSTTT	G....QPI. .DTATMDITAY WYANLREQVR
atdebs04p	AELGTITAVR	GSVPLHSTVT	G....EVI. .DTSAMDASY WYRNLRRPVL

Fig 20

17/26

atave02a	QHTQTLTYHP	PHTPLITANT	PPDQLLTPHY	WTQQARNTVD
atave05a	QHTQTLTYHP	PHTPLITANT	PPDQLLTPHY	WTQQARNTVD
atave04a	QHTQTLTYHP	PHTPLITANT	PPDQLLTPHY	WTQQARNTVD
atave08a	QHTQTLTYHP	PHTPLITANT	PPDQLLTPHY	WTQQARNTVD
atave03a	QHTQTLTYHP	PHTPLITANT	PPDQLLTPHY	WTQQARNTVD
atrap02a	AVAEGLYT	PQVA.....MA	AGDQVMTAEY	WVRQVRDTR
atrap11a	AVAEGLYT	PQVS.....MA	VGQVTTAEY	WVRQVRDTR
atrap08a	AVAEGLYT	PQVS.....MA	AGDQLTTTEY	WVRQVRDTR
atrap12a	AVAEGLYT	PQVS.....MA	VGQVTTAEY	WVRQVRDTR
atrap05a	TVAERLTYQT	PRLA.....MA	AGDRVTTAEY	WVRQVRDTR
atrap09a	AVAQGLTYHA	PGVV.....MA	AGDRVMTAEY	WVRQVRDTR
atfkb03a	DVAERLTYHE	PKLP.....MA	AGADCATPEY	WVRQVRDTR
atfkb07x	EAASGLTYHQ	PHT.....A	IPEDPTTAA	WARQVRDQVR
atfkb08x	ATTRELYDR	PHT.....A	IPNDPTTAEY	WAEQVRNPVL
atnid01a	DTLNTLNYQP	PTIPLISNLT	GQIADPNHLCTPDY	WIDHARHTVR
atnid03a	DTLNTLNYQP	PTIPLISNLT	GQIADPNHLCTPDY	WIDHARHTVR
atnid02a	DTLNTLNYQP	PTIPLISNLT	GQIADPNHLCTPDY	WIDHARHTVR
atnid00a	DTLNTLNYQP	PTIPLISNLT	GQIADPNHLCTPDY	WIDHARHTVR
atfkb10a	GVLESIAFGA	ARLPVVSTTT	GRDAAGD.LATPEH	WLRHARRPVL
atrap14a	TALESKFRFA	PALPVVSTVT	GRLIDQDEMGTPEY	WLRQVRRPVR
atmon06a	EAIRGVKFRQ	PSIPLMSNVS	GERA.....	.GEEITDPEY	WARHVRNAV
atmon08a	EAIREVKFTR	PKVSLISNVS	GLEA.....	.GEEIASPEY	WARHVRQTVL
atmon09a	EAIRGVKFRQ	PSIPLMSNVS	GERA.....	.GEEITSPEY	WARHVRQTVL
atepo02a	RVAASVTYRR	PSVSLVSNLS	GKVV.TDELSAPGY	WVRHVREAVR
atepo03x	RVAASVTYRR	PSVSLVSNLS	GKVVA.TDELSAPGY	WVRHVREAVR
atepo08a	RVTESVTYRR	PSIALVSNLS	GKPCT.DEVSAPGY	WVRHAREAVR
atepo00a	RVAESVSYRR	PSIVLVSNLS	GKACT.DEVSSPGY	WVRHAREVVR
atepo04a	RVAATIAYRA	PDRPVVSNVT	GHVAG.PEIATPEY	WVRHVSRAVR
atnid07a	AVAESVTYRT	PRPIVSEVT	GRPAAPSELMDPGY	WTRQIREPVR
attyl07a	EVAAGLRYRE	PELTVVSTVT	GRPARPGELTGPDY	WVAQVREPVR
atsor02a	RVAQSLTYHP	ARIPISNVT	GARATDHELASPDY	WVRHVRHTVR
atsorbl1a	RVAQGLTFHP	ARIPISNVT	GARATDQELASPET	WVRHVRDTR
atnys09a	AVVEDLTLQP	PLLPVVSNLT	GKPATVAQLTSADY	WVDHVRHVR
atnys12a	AVARGLTYHP	PTIPFVSNVS	GGLATAEQVRTPDY	WVGHVRAVR
atnys16a	AVAEGLEYHQ	PRIPVVSNT	GEVAAAEELCAADY	WVRHVRAVR
atnys17a	AVAEGLEYHQ	PRIPVVSNT	GEVAAAEELCAADY	WVRHVRAVR
atnys03a	AVAAGLTYHE	PRIPVLSNLT	GTVAAVADLCSADY	WVRHVREAVR
atnys15a	AVAEGLSYGE	PQIPVVSNT	GAVADGTLGTADY	WVRHVREAVR
atnys07a	VVAEGLSYAA	PSLPVVSNT	GQVATADELCSAEY	WVRHVREAVR
atnys08a	AVAEGLSYAT	PSLPVVSNT	GWLATADELCSAEY	WVRHVREAVR
atnys05a	AVAEGLSYAT	PTLPVVSNT	GRLATADDLCSAEY	WARHVREAVR
atnys06a	AVAEGLSYAT	PTLPVVSNT	GQVATADELCSAEY	WVRHVREAVR
atnys04a	AVVESLTFTA	PTTPVVSNT	GELAPAEALCSADY	WVRHVREAVR
atnys14a	TVAEGLEYHP	PRIPVVSNT	GDVADAADLCSADY	WVRHVREAVR
atnys00a	DVVSRLTFHQ	PSIPLVSNLT	GELA.GSEITSAEY	WVRHVREAVR
atnys10a	IVAEGLYTYRA	PRIPVSDLT	GRRADDAEVCTAEY	WVRHVREAVR
atnys18a	TALAPLTFAE	PEIPVVSNT	GLPATAEELATPHY	WVCHVRQAVR
atnys13a	RVAEGLYTYHE	PRIPVSTLL	GAPAGA.ELRTPDY	WVRHVREAVR
atave10a	RAAEQVTFSA	PRIPVVSNT	GAPLPAETMCTPDY	WVEHARSTVR
atrif02a	AVAERLTYRA	GSLPVVSTLT	GELAA...LDSPDY	WVGQVRNAV
atmon03a	EVARGLTFHA	PTLPVVSNT	GRLADAELMADA	WVRHVRRPVR
atave12a	QVAASLTYS	PAIPMVSTLT	GDIVAAGELSDPEY	WVRQVRRTVR
atrif09a	AELAGVTWRE	PEIPVVSNT	GRFAEPGELTEPGY	WAEHVRRPVR
atmon00a	AVAAGLTFHE	PVIPVVSNT	GELVTATATG	SGAGQADPEY	WARHAREPVR
attyl03a	RVARTLTFAP	PTIPLVSTLT	GTPVTEETLCTADH	WVRQAREPVR

Fig 2p

18/26

	401				450
atave00x	FEPAAARLLLQ	QGP.KTFVEM	SPHPVLTMGL	QELAPDLG..	DTTG
atdebs00p	FDEAIRSALE	VGP.GTFVEA	SPHPVLAAAL	QOTL.....	DAEG
atepo06p	FAAAAQALLE	GGP.ALFIEI	SPHPILVPPL	DEIQT.....	AE
atepo07p	FAAAAQALLE	GGP.ALFIEI	SPHPILVPPL	DEIQT.....	AE
atepo01p	FAEVVQAQLQ	GGH.GLFVEM	SPHPILTTSV	EEMRRA....	AQ
atepo05p	FAEAVQSLME	DGH.GLFVEM	SPHPILTTSV	EEIRRA....	TK
atsoralx	FSTAIHLLQ	EGF.DIFLEL	TPHPLALPAI	ESNLRR....	SG
atfkb01p	MDTAVSELDGSLFIEC	SAHPVLLPAL	DQ.....	
atfkb09p	MDTAIGELDGSLFIEC	SAHPVLLPAL	DQ.....	
atrap03p	FHPAVGQLQA	QGD.TVFVEV	SASPVLLQAM	DD.....	
atrap06p	FHPAVGQLQA	EGD.TVFVEV	SASPVLLQAM	DD.....	
atrap04p	FHPAVSQLQA	QGD.AVFVEV	SASPVLLQAM	DD.....	
atrap13p	FHPAVSQLQA	QGD.TVFVEV	SASPVLLQAM	DD.....	
atrap01p	FEPAAAGQLQA	QGD.TVFVEV	SASPVLLQAM	DD.....	
atrap07p	FDSAVGQLRA	EGD.TVFVEV	SASPVLLQAM	DD.....	
atrap10p	FHPAVSQLQA	QGD.TVFVEV	SASPVLMQAM	DD.....	
atfkb04x	FAAAVAAARE	PGD.TVFIEV	SAHPVLLPAI	NG.....	
attyl04p	FSAVVGGLLE	EGH.RRFIEV	SAHPVLVHAI	EQT....A..	EAAD
attyl06p	FSAVVGGLLE	EGH.RRFIEV	SAHPVLVHAI	EQT....A..	EAAD
attyl01p	FSAVVGGLLE	EGH.RRFIEV	SAHPVLVHAI	EQT....A..	EAAD
attyl02p	FSAVVGGLLE	QGH.RRFIEV	SAHPVLVHAI	EQT....A..	EAAD
attyl00p	FSAVVGGLLE	EGH.RRFIEV	SAHPVLVHAI	EQT....A..	EAAD
atnid05b	FERATRALIA	DGH.DVFLET	SPHPMLAVAL	EQT....V..	TDAG
attyl05b	FEKAVRALIA	DGY.DLFLEC	NHPMLAMSL	DET....L..	TDSG
atnid06x	FDATVRALLR	AGH.HTEIEV	GPHLLNAAI	DEI....A..	ADEG
atdebs01p	FADAVRALAE	QGY.RTFLEV	SAHPILTAAI	EEI....G..	DGSG
atmon02p	FADTIEALLA	DGY.RLFIEA	SAHPVLGLGM	EETIEQ....	AD
atmon10p	FADTIEALLA	DGY.RLFIEA	SAHPVLGLGM	EETIEQ....	AD
atmon04p	FADTIEALLA	DGY.RLFIEA	SAHPVLGLGM	EETIEQ....	AD
atmon07p	FADTIDALLA	DGY.RLFIEA	SAHPVLGLGM	EETIEQ....	AD
atmon11p	FADTIEALLA	DGY.RLFIEA	SPHPVLNLGI	QETIEQQA..	GAA
atmon12p	FADTIEALLA	DGY.RLFIEA	SPHPVLNLGM	EETIER....	AD
atmon05b	FADTIDALLA	DGY.RLFIEV	SPHPVLNLAL	EGLIER....	AA
atmon01p	FADAVTALLA	DGH.RVFIEA	SSHPVLTGL	QETFEE....	AG
atdebs02p	FHSAVQALTD	QGY.ATFIEV	SPHPVLASSV	QETL.....	DDAE
atdebs06p	FDEAVSAQSP	DGH.ATFVEM	SPHPVLTAAV	QE.....	IA
atave01p	FSDAVQALAD	DGH.RVFVEV	SPHPTLVPAI	EDTTEDTA..	ED..
atave07p	FSDAVQALAD	DGH.RVFVEV	SPHPTLVPAI	EDTTEDTA..	ED..
atave06p	FSHAIQTLTD	DGH.RAFIEI	SPHPTLVPAI	EDTTENTT..	EN..
atave09p	FSHAIQTLTD	DGH.RPFIEI	SPHPTLVPAI	EDTTENTT..	EN..
atnys01p	FADAVADLLA	AEY.RAFVEV	SSHPVLTMAV	LD....LI..	EEAG
atnys11p	FADAVADLLA	AEY.RAFVEV	SSHPVLMAV	QE....AI..	DEAG
atrif05p	FGPAVAELIE	QGH.GVFVEV	SAHPVLVQPI	SE....LT..	D..
atrif07p	FGPAVAELLE	QGH.GVFVEV	SAHPVLVQPI	SE....LT..	D..
atrif08p	FGPAVAELLG	LGH.RVFVEV	SAHPVLVQAI	SA....IA..	DD..
atrif10p	FGPAVEALLA	QGH.GVFVEL	SAHPVLVQPI	TE....LT..	DE..
atrif03p	FGPSVADLAG	LGH.TVFVEI	SAHPVLVQPL	SE....IS..	DD..
atrif06p	FGPAVAELVR	QGH.GVFVEV	SAHPVLVQPL	SE....LS..	DD..
atrif04p	FGPAVAELIE	QGH.RVFVEV	SAHPVLVQPI	NE....LV..	DD..
atrif01p	FGAAATALLE	QGH.TVFVEV	SAHPVTVQPL	SE....LT..	GD..
atnys02p	MEEATRALLA	AGH.RVFIEV	SPHPVLAAP	QETQEAVA..	EATG
atfkb02p	FGAAAAARLAE	LGH.RVFVEA	SPHPVLTAL	ADTLAG....	H
atave11p	FRDATQALVR	AGH.TVFIEA	CPHPAVAVGV	QETLDE.M..	GD
atdebs03p	FADAVTRLAE	SGY.DAFIEV	SPHPVVVQAV	EEAVEE.A..	DGAE
atnid04p	FESALRAMLA	DGV.DAFVEC	SPHPVLTVPV	ROTLED.A..	GA.
atdebs05p	FQDATRQLAE	AGF.DAFVEV	SPHPVLTVGI	EATLDS.A..	LPAD
atdebs04p	FEQAVRGLVE	QGF.DTFVEV	SPHPVLLMAV	EET....A..	EHAG

Fig 2q

19/26

atave02a	YATTTQTLHQ	HG.VTTYIEL	GPDNTLTTLT	HHNLPNPPTT	TLTLTHPHHH
atave05a	YATTTQTLHQ	HG.VTTYIEL	GPDNTLTTLT	HHNLPNTPTT	TLTLTHPHHH
atave04a	YATTTQTLHQ	HG.VTTYIEL	GPDNTLTTLT	HHNLPNTPTT	TLTLTHPHHH
atave08a	IATTTQTLHQ	HG.VTTYIEL	GPDNTLTTLT	HHNLPNTPTT	TLTLTHPHHH
atave03a	YATTTQTLHQ	HG.VTTYIEL	GPDNTLTTLT	HDNLPNTPTT	TLTLTHPHHH
atrap02a	FGEQVASFED	A...VFVEL	GADRSLARLV	DG.....
atrap11a	FGEQVASYED	A...VFVEL	GADRSLARLV	DG.....
atrap08a	FGEQVASYED	A...VFVEL	GADRSLARLV	DG.....
atrap12a	FGEQVASYED	A...VFVEL	GADRSLARLV	DG.....
atrap05a	FGEQVASYED	A...VFIEL	GADRSLARLV	DG.....
atnid02a	FGEQVASYED	A...VFVEL	GADRSLARLV	DG.....
atfkb03a	FAEQVAAYDG	A...ALLEI	GPDRLNARLV	DG.....
atfkb07x	FQAHAEYRPG	A...TFLEI	GPNDLSPVV	DG.....
atfkb08x	FHAHTQRYPD	A...VFVEI	GPGQDLSPLV	DG.....
atnid01a	FADAVQTAHD	QR.TTTYLEI	GAHPTLTTLT	HHTLDNP...
atnid03a	FADAVQTAHH	QG.TTTYLEI	GPHPTLTTLT	HHTLDNP...
atnid02a	FADAVQTAHD	QR.TTTYLEI	GPHPTLTTLT	HHTLDNP...
atnid00a	FADAVQTAHH	QG.TTTYLEI	GPHPTLTTLT	HHTLDNP...
atfkb10a	YADAVRELAD	LG.VNMFVAV	GPSGALASAA	SENTGGSAGT	YH.....
atrap14a	FQDAVRELAE	QG.VGTFVEV	GPSGALASAG	VECLGGDA.S	FH.....
atmon06a	FQPAIAQVAD	S..AGVFVEL	GPAPVLTTAA	QHTLDE.SD.	.SQES.....
atmon08a	FQPGIAQVAS	T..AGVFVEL	GPGPVLTTAA	QHTLDDVTD	HGPEP.....
atmon09a	FQPGVAQVAA	E..ARAFVEL	GPGPVLTTAA	QHTLDHITP	EGPEP.....
atepo02a	FADGVKALHE	AG.AGTFVEV	GPKPTLLGLL	PACLPEAEP.
atepo03x	FADGVKALHE	AG.AGTFVEV	GPKPTLLGLL	PACLPEAEP.
atepo08a	FADGVKALHA	AG.AGLFVEV	GPKPTLLGLV	PACLPDARP.
atepo00a	FADGVKALHA	AG.AGTFVEV	GPKSTLLGLV	PACMPDARP.
atepo04a	FGDGAKALHA	AG.AATFVEV	GPKPVLGLL	PACLGEADA.
atnid07a	FAAAVRAARA	AG.AATFVEL	GPDAVLGMA	RECAAG....DTGT
atty107a	FADAVRTAHR	LG.ARTFLET	GPDGVLGMA	EECLD....DTVA
atsor02a	FLDGVRALHA	EG.ARVFLEL	GPHAVLSALA	QDALGQ....D.EGTS
atsorbl1a	FLDGVRTLHA	EG.ARAFLEL	GPHPVLSALA	QDALGH....D.EGPS
atnys09a	FADGIDWLA.	RHDTTAFLEL	GPDGVLSAMA	QDCLDA....A.DAD.
atnys12a	FADGIDWLAT	QGDVHTFLEL	GPDGVLSAMA	RESLTD....P.SRT.
atnys16a	FADGVRTLAE	RG.ATAFLEI	GPDGVLSALA	RGVL.....P.AEA.
atnys17a	FADGVRTLAE	RG.ATAFLEI	GPDGVLSALA	AACL.F....D.TDA.
atnys03a	FADGVTALTD	RG.VTTLVEL	GPDGVLSAMA	QESI.....P.DGA.
atnys15a	FADGIRALTD	AG.VGAFLEL	GPDGTLAALA	QQA.....P.D.A.
atnys07a	FADGVTALEA	EG.VRTFLEL	GPDGVLAAAMA	GASL.....T.ESS.
atnys08a	FADGITTLEA	EG.VRTFLEL	GPDGILSALA	QQL.....A.GEA.
atnys05a	FADGVSTLEN	EG.VTTFLEL	GPDGVLSAMA	QQL.....T.GDA.
atnys06a	FADGVTALEA	EG.VRTFLEL	GPDGVLAAAMA	RETV.....A.DDT.
atnys04a	FADGIRTLAD	RG.VTTFVEL	GPDSVLSAMA	QESA.....P.EGA.
atnys14a	FADGVRTMAD	RG.VHLFLEL	GPDAVLSAMA	RQCA.....P.D.A.
atnys00a	FADGITALAK	AG.ADVLEL	GPGGVLSAMA	RDTL.G....P.DST.
atnys10a	FADCVRTLRD	AG.ATTFLEL	GSDGLLTAMA	EDTL.G....D.DHD.
atnys18a	FGDGVRALAD	RG.VRTFLEL	GPDGVLSALV	RENL.....P.EPG.
atnys13a	FADGVRALHD	AG.AGTFVEI	GPDGVLTAALT	QQTLDT....V.EAGA
atavel0a	FADGISWLQE	QG.VTTCLEI	GPDGTLAALA	QDSLISA....P.....
atrif02a	FSDAVTALGA	QG.ASTFLEL	GPGGALAAMA	LGTLGG....P.EQSC
atmon03a	FHDGLRALSE	QGVVR.YLEL	GPDPVLATMV	QDGLPA....P.AEGE
atavel2a	FGDAISRLHT	DG.VRTFMEL	GPDGTLAALA	EECLEATADS	HPADD.DTGT
atrif09a	FAEGVAAATE	SGG.SLFVEL	GPGAALTALV	EET.....
atmon00a	FLSGVRGLCE	RG.VTTFVEL	GPDAPLSAMA	RDCFPAPADR	SRPRP.....
atty103a	FLDAMRTLRA	DG.IDTFVEL	GPDGVLSAMA	RDCADDRPDG	DTTGAGDGET

Fig 2r

20/26

451 500

atave00x TADTVIMGTL RRGQGTLDHF LTSLAQLRGH GE..TSATTV LSARLTALSP
atdebs00p SSAAVV.PTL QRGQGGMRRF LLAAQAFTG GV..AVDWT AAYDDVGA.EP
atepo06p QGGAHV.GSL RRGQDERATL LEALGTLWAS G..YPVSWAR LFPAGG....
atepo07p QGGAHV.GSL RRGQDERATL LEALGTLWAS G..YPVSWAR LFPAGG....
atepo01p RAGAAV.GSL RRGQDERPAM LEALGTLWAQ G..YPVPWGR LFPAGG....
atepo05p REGVAV.GSL RRGQDERLSM LEALGALWVH G..QAVGWER LFSAGGAGL.
atsoralx RRGVVL.PSL RRNEDERGVM LDTLGVLVVR G..APVRWDN VYPA...AF.
atfkb01p .E.RTV.ASL RTDDGGWDRF LAALAQAFTQ GA..DVDWTT LIEPA.....
atfkb09p .E.RTV.ASL RTDDGGWDRF LTALAQAFTQ GA..DVDWTT LIAPA.....
atrap03p .DVVTV.ATL RRDDGDATRM LTALAQAFTQ GV..TVDWPA ILG.T.....
atrap06p .DVVTV.ATL RRDDGDATRM LTALAQAFTQ GV..TVDWPA ILG.T.....
atrap04p .DVVTV.ATL RRDDGDATRM LTALAQAFTQ GV..TVDWPA ILG.T.....
atrap13p .DVVTV.ATL RRDDGDATRM LTALAQAFTQ GV..TVDWPA ILG.T.....
atrap01p .DVVTV.ATL RRDDGDATRM LTALAQAFTQ GV..TVDWPA ILG.T.....
atrap07p .DVVTV.ATL RRDDGDATRM LTALAQAFTQ GV..TVDWPA ILG.T.....
atrap10p .DVVTV.ATL RRDDGDATRM LTALAQAFTQ GV..TVDWPA ILG.T.....
atfkb04x ...TTV.GTL RR.GGGADRV LDSLAKAHTV GV..AVDWST VVAATGAADD
attyl04p RSVHAT.GTL RRQDDSPHRL LTSTAEAWAH G..ATLTW..
attyl06p RSVHAT.GTL RRQDDSPHRL LTSTAEAWAH G..ATLTW..
attyl01p RSVHAT.GTL RRQDDSPHRL LTSTAEAWAH G..ATLTW..
attyl02p RSVHAT.GTL RRQDDSPHRL LTSTAEAWAH G..ATLTW..
attyl00p RSVHAT.GTL RRQDDSPHRL LTSTAEAWAH G..ATLTW..
atnid05b TDAAVL.GTL RRRHGGPRAL ALAVCRAFAH GVE..VDPEA VF.....
attyl05b GHGTVM.HTL RRQKGSADKF GMALCLAYVN GLE..IDGEA LF.....
atnid06x VAATAL.HTL QRGAGGLDRV RNAVGAAFAH GVR..VDWNA LF.....
atdebs01p ADLSAI.HSL RRGDGSGLADF GEALSRAFAA GVA..VDWES VH.....
atmon02p MPATVV.PTL RRDHGDTTQL TRAAHAFTA G..ADVDRR WF.....
atmon10p IPATVV.PTL RRDHGDTTQL TRAAHAFTA G..APVDWRR WF.....
atmon04p IPATVV.PTL RRDHGDTTQL TRAAHAFTA G..ADVDRR WF.....
atmon07p IPATVV.PTL RRDHGDTTQL TRAAHAFTA G..ATVDWRR WF.....
atmon11p GTAVTI.PTL RRDHGDTTQL TRAAHAFTA G..APVDWRR WF.....
atmon12p MPATVV.PTL RRDHGDAAQL TRAAQAFTA G..AEVDWTG WF.....
atmon05b VPATVV.PTL RRDHGDTTQL TRAAHAFTA G..ADVDRR WF.....
atmon01p VDAVTV.PTL RREDGGARL ARSLAQAFTA G..CAVRWEN WF.....
atdebs02p SDAAVL.GTL ERDAGDADF LTALADAHTR GVA..VDWEA VL.....
atdebs06p ADAVAI.GSL HRDTAE.EHL IAEARAHVH GVA..VDWRN VF.....
atave01p ..VTAI.GSL RRGDNDTRRF LTALAHHTT GIGTPTTWHH HY.....
atave07p ..VTAI.GSL RRGDNDTRRF LTALAHHTT GIGTPTTWHH HY.....
atave06p ..ITAT.GSL RRGDNDTRRF LTALAHHTT GIGTPTTWHH HY.....
atave09p ..ITAT.GSL RRGDNDTRRF LTALAHHTT GIGTPTTWHH HY.....
atnys01p VTAVAT.GTL RRDQGGAGRF LLSAAEVFVR GV..DVDWAG AF.....
atnys11p VPAVAA.GTL RRDQGGTDRF LLSAAEVFVR GV..DVDWAG LF.....
atrif05p ..AVVT.GTL RRDDGGVRR LLSMAELFVR GV..PVDWAT MA.....
atrif07p ..AVVT.GTL RRDDGGVRR LLSMAELFVR GV..PVDWAT LV.....
atrif08p TDAVVT.GSL RREDGGLRRL LLSMAELFVR GV..DVDWAT MV.....
atrif10p TAAVVT.GSL RRDDGGLRRL LLSMAELFVR GV..EVDWTS LV.....
atrif03p ..AVVT.GSL RRDDGGLRRL LLSAAELYVR GV..AVDWT AAV.....
atrif06p ..AVVT.GSL RREDGGLRRL LLSMAELYVQ GV..PLDWT A VL.....
atrif04p TEAVVT.GTL RREDGGLRRL LLSMAELFVR GV..TVDWSG VL.....
atrif01pAI.GTL RREDGGLRRL LLSMAELFVR GI..DVDWT A MV.....
atnys02p GSAVVL.GSL RREDGGPRRF LTSLEAHTH GA..PVDWT T TF.....
atfkb02p PNTAVT.GTL RRGDGGARRF TRSLAELWVR GV..PVS...
atavellp LDSLVV.GSL RRGEGGLRRL LMSVAELFVG GV..AVEWSG VF.....
atdebs03p .DAVVV.GSL HRDGGDLASF LRSMATAHVS GV..DIRWDV AL.....
atnid04p .GAVAV.GSL RRDDGGLRRL LTSAAEAQVA GV..PVDWAA LC.....
atdebs05p AGACVV.GTL RRDRGGLADF HTALGEAYAQ GV..EVDWSP AF.....
atdebs04p AEVTCV.PTL RREQSGPHEF LRNLLRAHVH GVGADL....

Fig 2s

21/26

```

atave02a PQTH.....LLTNL AK.....TT T..TWHPPHY
atave05a PQTH.....LLTNL AK.....TT T..TWHPPHY
atave04a PQTH.....LLTNL AK.....TT T..TWHPPHY
atave08a PQTH.....LLTNL AK.....TT T..TWHPPHY
atave03a PQTH.....LLTNL AK.....TT T..TWHPPHY
atrap02a .....IAML HGD.HE....AQAAVGAL AHLYVNG.VS V..EW.SAVL
atrap11a .....VAML HGD.HE....AQAAVGAL AHLYVNG.VS V..EW.SAVL
atrap08a .....VAML HGD.HE....AQAASVAL AHLYVNG.VT V..DW.PALL
atrap12a .....VAML HGD.HE....IQAAIGAL AHLYVNG.VT V..DW.PALL
atrap05a .....VAML HTD.HE....AQAAISAL AHLYVNG.VT V..DW.TALL
atrap09a .....VAML HGD.HE....TQAAIGAL AHLYVNG.VT V..DW.TALL
atfkb03a .....IPVL HGE.DE....ARSAMTAL ARLHTGG.VA V..DW.PEVI
atfkb07x .....IPTQ TGTPEE....VQALHTAL ARLHTRG.GV V..DW.PTVL
atfkb08x .....IALQ NGTAE....VHALHTAL ARLHTRG.AT L..DW.SRIL
atnid01a ....TTIPTL HREHPEPETL TTAL....AT ..LHTTGHTT T.....
atnid03a ....TTIPTL HREHPEPETL TTAL....AT ..LHTTGHTT T.....
atnid02a ....TTIPTL HREHPEPETL TTAL....AT ..LHTTGHTT T.....
atnid00a ....TTIPTL HRERPEPETL TQAI....AA VGVRTDGIDW A.....
atfkb10a .....AVL RARTGEES...AALTAV AELHAGG.AP V..DL.AAVL
atrap14a .....AVL RPRSPEDV...CLMTAI AELHAGG.TA I..DW.AKVL
atmon06a ....VLVASL AGERPEES...AFVEAM ARLHTAG.VA V..DW.SVLF
atmon08a ....VLVSSL AGERPEES...AFVEAM ARLHTAG.VA V..DW.SVLF
atmon09a ....VVTASL HPDRPDDV...AFAHAM ADLHVAG.IS V..DW.SAYF
atepo02a ....TLLASL RAGREEA...AGVLEAL GRLWAAGGS. V..SW.PGVF
atepo03x ....TLLASL RAGREEA...AGVLEAL GRLWAAGGS. V..SW.PGVF
atepo08a ....VLLPAS RAGRDEA...ASALEAL GGFVVVGGGS. V..TW.SGVF
atepo00a ....ALLASS RAGRDEP...ATVLEAL GGLWAVGGL. V..SW.AGLF
atepo04a ....VLVPSL RADRSEC...EVVLAAL GAWYAWGGA. L..DW.KGVF
atnid07a AFAAALRRGR ....PEC...ATVLPAA ATAFVQG.AH V..DW.AAPY
atty107a LLPAIHKPGT APHGPA...PGALRAA AAAYGRG.AR V..DW.AGMH
atsor02a PCAFL..PTL RKGRDDA...EAF TAL GALHAAG.LT P..DW.SAFF
atsorbl1 PCAFL..PTL RKGRDDA...EAF TAL GALHAAG.LT P..DW.NAFF
atnys09a .AVTL..PAL RAGRPEE...HTLTAL AGLHVHG.AT L..DW.TGCF
atnys12a .AL.L..PTL RGDRPEE...PALVTAV AAAHAGG.AR V..DW.SGYF
atnys16a .L.VT..PTL RKDRDEE...SALLAGL ARLHVAG.VT V..DW.SAAL
atnys17a .E.VV..PAL RAGRPEE...HTALTAA AQLHVAG.VD I..DW.TAVL
atnys03a .A.AV..PLL RKDRPEE...LSAVTGL ARAHVRG.VT V..RW.AGLF
atnys15a .V.SV..PVL RKDRDEE...PAAVAAL ARLHTAG.VP V..DW.TAFY
atnys07a .L.AV..PLL RKDRPEE...PAALAAL AQLHIAG.AR V..DW.PVLF
atnys08a .V.TV..PVL RKDRGEE...STALTAR ARLHTRG.LI E..DW.QDFF
atnys05a .A.TV..PAL RKDRDEE...TSALTAL ARLHTAG.LR V..DW.AAFF
atnys06a .V.TV..PVL RRMNPEE...RTLLTAL GRLHTTG.TP I..DW.AALL
atnys04a .G.TI..PLL RRDRPEE...QAVLAAL CHLQVLG.VE A..DW.SATF
atnys14a .V.VV..PAL RNRNDED...ETLVGAV ARLHVHG.AG P..RW.DAYF
atnys00a .TDVV..PAL SKGRPEE...TAFAGAL GRLHTLG.VP V..DW.PAFY
atnys10a .AELV..PML RAGRAEE...LAAATAL ARLQVRG.VD V..DW.AAYL
atnys18a .LVAV..PVL RKERPEE...TTVLAAL GTLWAGG.AD V..DW.DAVF
atnys13a PAVVV..PLQ RRDRAGD...LALLEGL ATLHTHG.TG P..SW.PAYF
atave10a .ARAI..PAL RPDQPEA...RSVMTAL AELFVAG.TA V..EW.AGVF
atrif02a ....V..ATL RKNGAEV...PDVLTAL AELHVRG.VG V..DW.TTVL
atmon03a EPEPVVAAAL RSKHDEG...RTLLGAV AALHTDG.QP A..DL.TALF
atave12a PQENLLIPLL RPDSPEP...GTLLTGL ARLHTHGAAA V..NW.PAAL
atrif09a .AEVTCVAAL RDRPEV...TALITAV AELFVRG.VA V..DW.PALL
atmon00a ....AAIATC RRGREDEV...ATFLRSL AQAYVRG.AD V..DF.TRAY
atty103a PDPLLTPLLL RRSVPETGDA EHPGGFERAL ATAYAHGV...PLRL

```

Fig 2t

22/26

	501				550
atave00x	TQQQSLLLDL	VRAHTMAVLN	DDGN~~~~~	~~~~~	~~~~~
atdebs00p	GSLPE.FAPA	EEEDEPAESG	VDWNAPPHVL	RER~~~~~	~~~~~
atepo06p	RRVPLPTYPW	QHERCWIEVE	PDARR~~~~~	~~~~~
atepo07p	RRVPLPTYPW	QHERYWIEDS	VHGSKPSLRL	RQLRNGATDH
atepo01p	RRVPLPTYPW	QRERYWIEAP	AKSAAGDRRG	VRAGGHPLL
atepo05p	RRVPLPTYPW	QRERYWVDAP	TGGAAGGSRF	AHAGSHPLL~
atsoralx	ESMPLPSTAG	~~~~~	~~~~~	~~~~~
atfkb01pP.H	RVLDLPTYPF	DHKRYWLQPA	PVT~~~~~	~~~~~
atfkb09pP.D	RVLDLPTYPF	DHKRYWIEAT	GAADLTALGL	TDTAHP~~~~
atrap03pTTT	RVLDLPTYAF	QHORYWVE..	.GVDRSAAG.	...GHPLLGV
atrap06pATT	RVLDLPTYAF	QHORYWLR..	.SVDRAAAD.	...GHPLLGT
atrap04pTTA	RVLDLPTYAF	QHORYWVK..	.SVDRAAAD.	...GHPLLGA
atrap13pTTT	RVLDLPTYAF	QHORYWLK..	.SVDRAAAD.	...GHPLLGT
atrap01pTAA	RVLDLPTYAF	QHORYWLK..	.GVDRRAAD.	...GHPLLGT
atrap07pATT	RVPDLPTYAF	QHOREWAE..	.GADRSVAG.	...GHPLLGV
atrap10pPAT	RVLDLPTYAF	QHORYWAEAG	RSADVSAAGL	DAVGHPPLLGA
atfkb04x	AASVTAHDTG	TAHDLPTYAF	HHERYWIEPA	TGTDASGLGL	D~~~~~
attyl04p	...DPALPPG	HLTTLPTYPF	NHHHYWLDTT	PTTPA.TTTQ	SPTDAQNPAD
attyl06p	...DPALPPG	HLTTLPTYPF	NHHHYWLDTT	PTTPA.TTTQ	SPTDAWR...
attyl01p	...DPALPPG	HLTTLPTYPF	NHHHYWLDTT	PTTPA.TTTQ	SPTDAWR...
attyl02p	...DPALPPG	HLTTLPTYPF	NHHHYWAVTS	PAGVG.DAA.AGR...
attyl00p	...DPALPPG	HLTTLPTYPF	NHHHYWLDTI	DGGGGDDATQ	EKESGPLTRE
atnid05b	.G.....PGA	RPVELPTYPF	QRERYWCHP.	GVRGGDPASL	GMDGADHPLL
attyl05b	.G.....PDS	RRVNPPTYPF	QRERYWYHPT	SGRRGDITAA	GVAEAEHPLL
atnid06x	EG.....TGA	RRVPLPSYAF	HRDRFWLPTA	AARRPATSSS	~~~~~
atdebs01p	LG.....TGA	RRVPLPTYPF	QRERVWLEPK	PVARRSTEVD	EV~~~~~
atmon02pPADPAP	RTIDLPTYAF	QRRRYWLADT	VKRDSGWDPA	GS~~~~~
atmon10pPADPTP	RTVDLPTYAF	QHORYWLEERS	ASASGAVSGE	QSA~~~~~
atmon04pPADPTP	RTVDLPTYAF	QHORYWLEEP	SGLTGDAADL	GMVA~~~~~
atmon07pPADPTP	RTIDLPTYAF	QRRSYWL..P	VDGVGDVRS	GLRRVE~~~~
atmon11pPADPTP	RTVDLPTYAF	QHKHYWVEPP	AAVAAGVGGH	DPVEA~~~~~
atmon12pPAVPLP	RVVDLPTYAF	QRERFWLEGR	RGLAGDPAGL	GL~~~~~
atmon05bPADPAP	RTVDLPTYAF	QRQDFWPAPA	GGRSGDPAGL	GLAASGHP~~
atmon01pPATGT.	STVELPTYAF	QRRRYWLEAP	TG.TQDAAGL	GL~~~~~
atdebs02pGRA	GLVDLPGYPF	QGKRFWLLPD	RTTPRDEL.D	GMVA~~~~~
atdebs06pPAA	PPVALPNYPF	EPQRYWLAPE	VS...DQLAD	SRYRVD~~~~
atave01p	THHHTHPHPH	THLDLPTYPF	QHORYWLESS	QPGAGSGSG~	~~~~~
atave07p	THHHTHPHNH	.HLDLPTYPF	QRQHYWLD.A	PTGAGDV~~~	~~~~~
atave06p	TQTHPHPNPH	THLDLPTYPF	QHORYWLQPP	TTTTDLTTTG	LTPTHHPL~~
atave09p	TQTHPHPNNH	.HLDLPTYPF	QHORYWLQ~~	~~~~~	~~~~~
atnys01p	E.....GTGA	ARVDLPTYAF	QRERYW.NTR	TAADRTPAD	PMDAEFWA~~
atnys11p	E.....GTGA	SRIDLPTYAF	QHEHLW.AVP	PAPEAVAAAD	PDDAAFWTAV
atrif05pPPA	.RVELPTYAF	DHQHFW..LS	PPAVA.DAPA	LGLAGADHPL
atrif07pPPA	.RVDLPTYAF	DHQHFW..LR	PAAQA.DAVS	LGQAAAEHPL
atrif08pPPA	.RVDLPTYAF	DHQHYW..LR	YVETATDAA~	~~~~~
atrif10pPPA	.RADLPTYAF	DHEHYW..LR	AADTASDAVS	LGLAGADHPL
atrif03pPAA	GWVDLPTYAF	DRRHFW..LH	EAETAEEAEG	M~~~~~
atrif06pPRT	GRVDLPKYAF	DHRHYW..LR	PAESATDAAS	LGQGAADHPL
atrif04pPPS	RRVELPTYAF	DHQHYW..LQ	MGGSATDAV~	~~~~~
atrif01pPAA	GWVDLPTYAF	EHRHYW..LE	PAEPASAGDP	LLGT~~~~~
atnys02p	A.....RSAY	QPVDLPTYPF	QRQDFWPEAR	PATPAAGADA	SD~~~~~
atfkb02p	P.....FGEL	RGVPLPTYPF	RRDRYWDVAE	PAGTSGHP~~	~~~~~
atave11p	GSVGRGVAGG	CGVELPTYAF	ERERFWLDVE	GAPRGSGVSG	QW~~~~~
atdebs03pPGA	APFALPTYPF	QRKRYWLQPA	APAAASDELA	YRV~~~~~
atnid04pPRA	GWVDLPTYAF	QRERYWVAPA	EPGPAAGAGS	AAATGPAAA~
atdebs05pADA	RPVELPVYPF	QRQRYWLPIP	TGGRARDEDD	DWR~~~~~
atdebs04p	...RPAVAGG	RPAELPTYPF	EHQRFWRPH	RPAVVSALGV	R~~~~~

Fig 2u

23/26

```

atave02a THHDNQPHTH THLDLPTYPF QHHHYWLE.. STQPGAGNV~ ~~~~~
atave05a THHHNQPHTH THLDLPTYPF QHHHYWLELP SAQTS PGQRR SRRSAPD~~~
atave04a THHHNQPHTH THLDLPTYPF QHQHYWLE.. STQPGAGSGS GSGSGRAG~~
atave08a THHHNQPHTH THLDLPTYPF QHHHYWLE.. STQPGAGNVS AA~~~~~
atave03a THHHNQPHTH THLDLPTYPF QHHHYWLQ.. ..PPGKPSDP SP~~~~~
atrap02a GDVPVTRV.. ..LDLPTYAF QHQRYWLE.. .GTDRATAG. ...GHPLLGS
atrap11a GDVPVTRV.. ..LDLPTYAF QHQRYWLE.. .GTDRATAG. ...GHPLLGS
atrap08a GDAPATRV.. ..LDLPTYAF QHQRYWLE.. .GTDRMAAG. ...GHPLLGE
atrap12a GDAPATRV.. ..LDLPTYAF QHQRYWLE.. .GTDRATAG. ...GHPLLGS
atrap05a GDAPATRV.. ..LDLPTYAF QHQRYWLE.. .GADRAAG. ...GHPLLGP
atrap09a GDVPVTRV.. ..LDLPTYAF QQQRYWAEVG RSADVSGAGL DAVGHPLLGA
atfkb03a GAAP.TDL.. ..PHLPTYPF ERTRYWLSR AAGDA~~~~~
atfkb07x .GSDRAPV.. ...ALPTYPF QHKDYWLRAT AQVDVTGAGQ EKVAHPLL~
atfkb08x GGASRHDP.. ...DVPSYAF QRRPYWIE.S APPATADSG. ....HPVLGT
atnid01a ..LHTTSPQT HHLDLPTYPF QRDYWM.EP VRVAQVSGQP GADRLRYRVV
atnid03a ..LHTTSPQS HHLDLPTYPF QRDYWM.AV PPRAAVGDLA ~~~~~
atnid02a ..PHPSHIPA QRVSLPAYPF QRRAYWM..P NSAAHIGRSD AEAATRLGLA
atnid00a ..VLCGASRP RRVELPTYAF QRRTHWAPGL TPNHAPADRP AAEPQRAMAV
atfkb10a A.....GG RPVDLPVYPF QHRSYWLAPA VGGGSPTAVP D~~~~~
atrap14a S.....GG RAVDLPVYPF QHQS YWLAPA ..APDATAVA PVVEEGGEY
atmon06a AGDRVPGL.. ..VELPTYAF QRERFWLSG. RSGGGDAATL GLVAAG~~~~
atmon08a AGDRVPGL.. ..VELPTYAF QRERFWLSG. RSGGGDAATL GLVAAGHPL~
atmon09a PDDPAPRT.. ..VDLPTYAF QGRRFWLADI AAPEAVSSTD GEEA~~~~~
atepo02a .....PTAG RRVPLPTYPW QRQRYWIEAP AE~~~~~
atepo03x .....PTAG RRVPLPTYPW QRQRYWPDIE PDSRR.HAAA DPTQGWFY~
atepo08a .....PSGG RRVPLPTYPW QRERYWIEAP VDREA.DGTG ~~~~~
atepo00a .....PSGG RRVPLPTYPW QRERYWIDTK ADDAA.RGDR RAPGAGHDEV
atepo04a .....PDGA RRVALPMYPW QRERHWMDLT PRSAA.PAGI AGRWPLAGVG
atnid07a ...EG..AGA RRVDLPTYPF QHTRYWL~~~
atty107a A..DGPEGPA RRVELPVHAF RHRRYWLAPG RAA~~~~~
atsor02a A.....PFAP R~~~~~
atsorbl1a A.....PFAP CKVPLPTYTF ~~~~~
atnys09a AGT.....GA RRTDLPTYAF QRRRYWPKAL QSGTA.DLRS VGLGAA~~~~
atnys12a ADH.....GA RRTTLPTYAF QRERYWPDIT AATSA.HTPG SALDAEFW~
atnys16a TGT.....GA RGTDLPTYAF QRERYWPE.. LAAEP.AG.. GGADAADA~
atnys17a AGT.....GA RRIALPTYAF QRERYWPS.. LAAQA.PGDA GG~~~~~
atnys03a DGT.....GA RRADLPTYPF QHQRFWPT.. AAR.A.AQDV TAAGLGAADH
atnys15a AGT.....GA HRTDLPTYAF QYERYWPK.. ATY.R.PADA TGL~~~~~
atnys07a AGV.....GA GRVELPTYAF QRGWFWPV.. GRVGV.GGDV ~~~~~
atnys08a AGV.....GA GRVELPTYAF QRGWFWPV.. GRVGV.GGDV GAVGLGSAGH
atnys05a AGS.....GA TRVDLPTYAF QHATYWPT.. GTLPT..AHA AAVGL~~~~
atnys06a APT.....GA RPVDLPTYAF QHRPFWPS.. GPRDT..ADA AAVGIAGASH
atnys04a RGL.....DP VRVDLPTYAF QHRWFPA.. ARPAR.PDDV RAAGLGAA~
atnys14a AGR.....GA QWLDLPTYPF QRGWFPE.. SLPGA.ASAA PAAGQPA~~
atnys00a AGT.....GA RRVELPTYAF QHVRHWPT.. PPRPN.GAGP GALGHPLL~
atnys10a AGT.....GA RRTDLPTYAF QHAYYWPQ.. LPTFA.AALA AADPADQQLW
atnys18a AGT..RTPQA DPVELPTYAF QRARYWPTLG ARHGD.PADL G~~~~~
atnys13a EAT.....GG HRTDLPTYAF QRERYWPELG APVAT.APQD PAAW~~~~~
atave10a EGTAREVGDG CGVELPTYAF ERERFWLDVE EGSAG.GSGV SCMWGGPLWE
atrif02a ....DEPATA VGTVLPTYAF QHQRFWVDVD ET~~~~~
atmon03a .....PADA GOVPLPTYRF QRRRYWRVAP DAAAP.ARAA GLQ~~~~~
atave12a PERDR....A RHLDLPTYAF DHHR YWVDT S AGHPG.DLSA AGLGT~~~~
atrif09a PPVTGF.... ..VDLPKYAF DQOHYWLQPA AQATD.AASL GOV~~~~~
atmon00a GAT.....AT RRFPLPTYPF QRERHWPAAA GVGQQ.PETP ELP~~~~~
atty103a APAPDAASLA VAAELPTYAF QRTHYWLDA AAPAALPAGL DDAGHPLLSA

```

LPTY motif

Fig2v

24/26

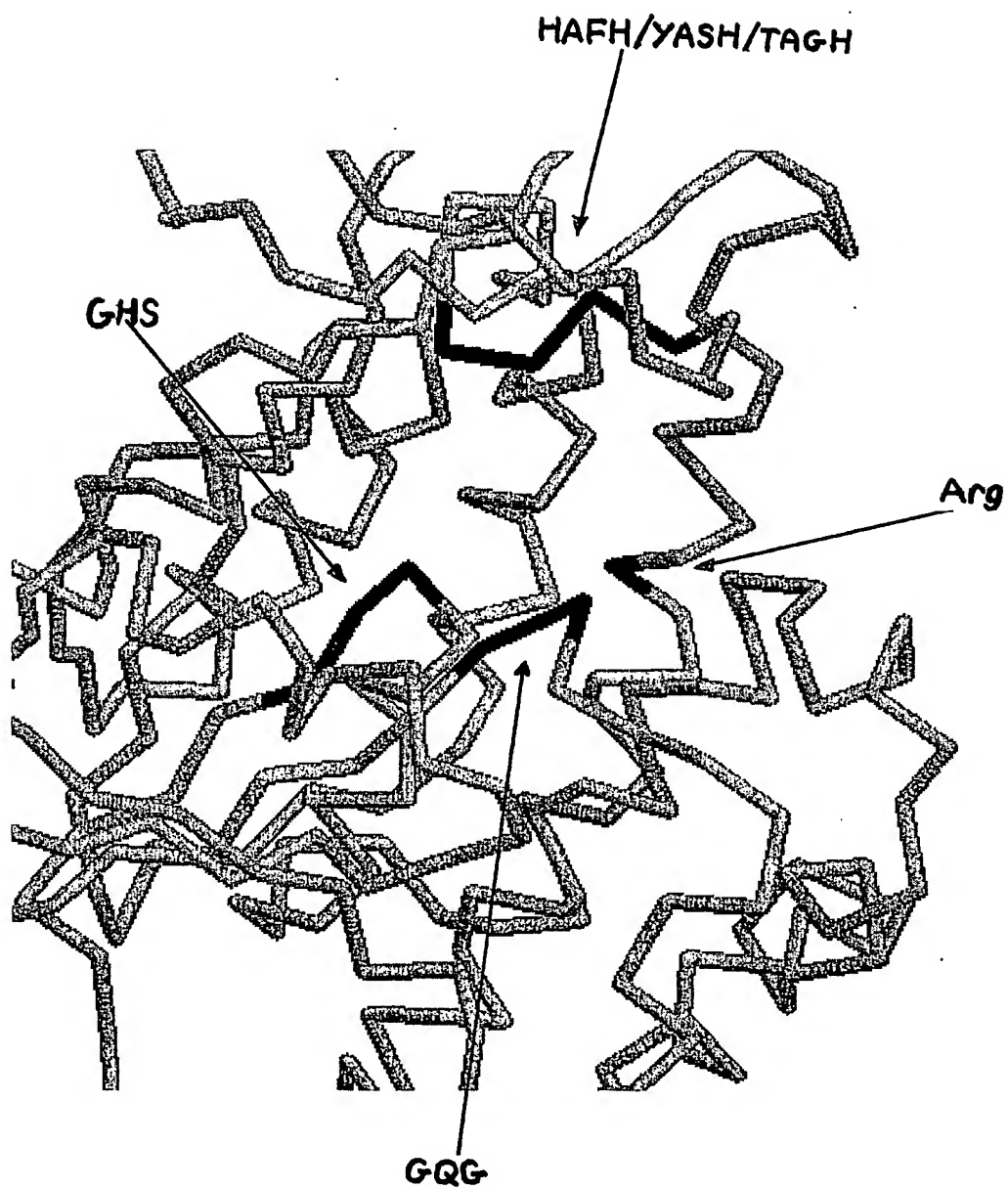


Fig3

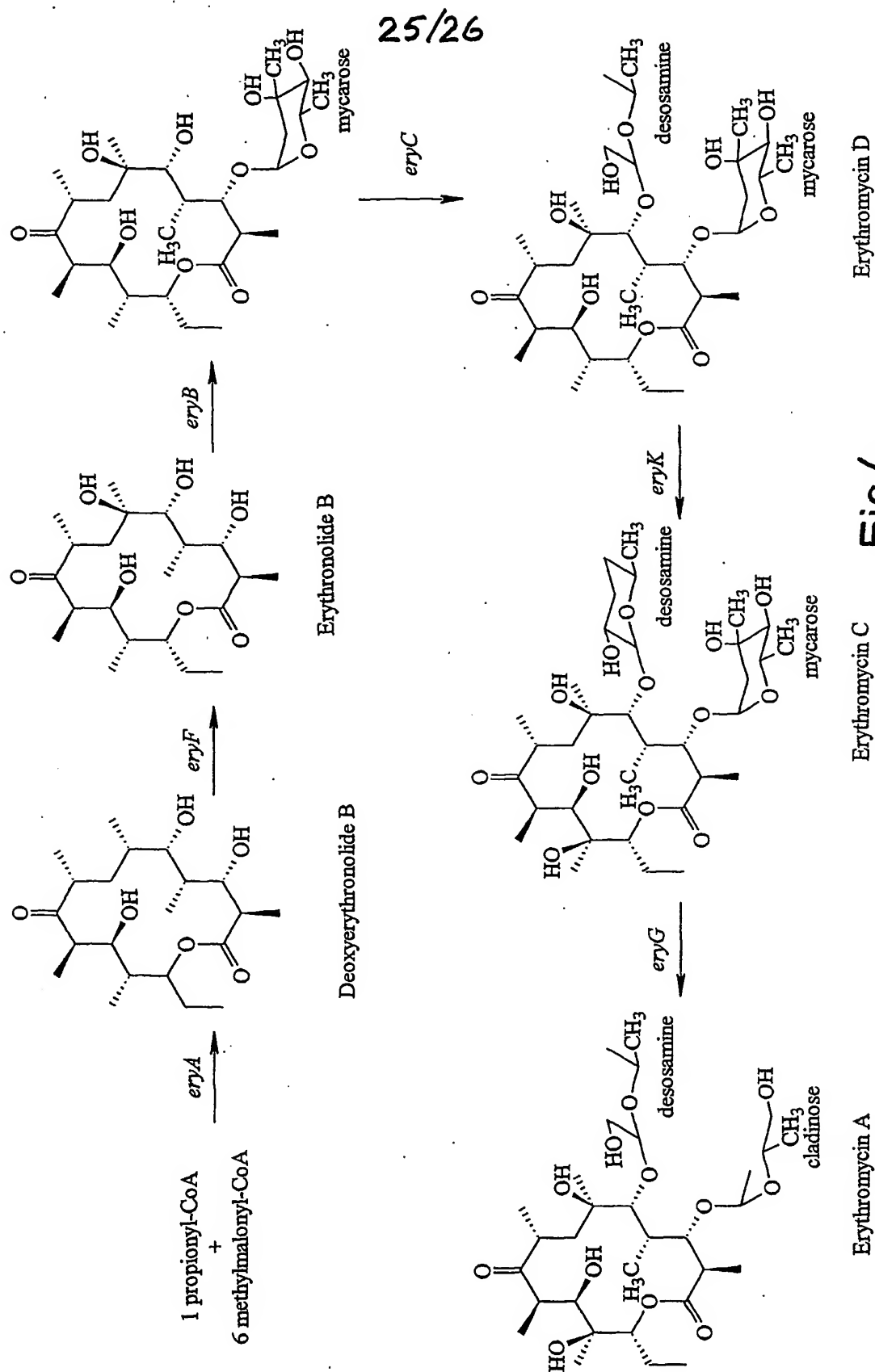


Fig4

26/26

15500 GGTGGTCCGACGCCATGGCCCGTGGCGCTCACAGCGCTTCGGCGCTGCGCGCGCAGGC 15559
 15560 CGGTGCGCTGGGACGACCTCGCCGCCACCGCCCAACCGCCCGACCGCGCGCGGTGCG 15619
 15620 CCACGGCTGCCACACCCGTGCGCCCTCGCCCAACCGCGCGGTCTGTCTGCGCGCGA 15679
 15680 CACCGCGAACTGTGGGTCTCTTGACGCGCTGGCGAGGCGCGAGACCGCGTCCAT 15739
 15740 CGTGGCGGCGAGGCGTACACCGAGGAGGACGCGCTTCCTCTTCACTGGGCGAGGAGC 15799
 15800 GCAACGCTCGGCATGGGCGGGAGTTGATCCGCTGTTCCCGCTTCGCCGACGCTCT 15859
 15860 CGACGAGGCTTCGCCGCCCTGGACGTACATCTGACCGCCACTGCGCGAGATCGTCTT 15919
 15920 GGGCAGACCGACTCGGGTGGAACTCTCGGTGAGATGTATGCGCGAGGTGCCGA 15979
 15980 CCATCAGGCACTCTCGACACAGACCGCTACACCAAGCCCGCTCTTCGATCGAGAC 16039
 16040 GAGCTGTACCGGTGGAGCTCTTCGGCTGAAGCCGACTACGTCTCGGCCACTC 16099
 16100 GGTGGCGAGATCGCCCGCGCACGTGCGCGGTCTCTGTTGCCGACCGGAGCGC 16159
 16160 TCTGGTGGCCACGCGGGGACGGCTCATGACGGGTTTCGCGCGCGGATGGCCG 16219
 16220 GTGGAGGCCACGGCGGACGAGGCGGCGAAGCTCGCTGCTCGCGGCGACCGCGTCA 16279
 16280 CGTGGCGCGCTCAACGGCCCGACTCCGTGCTCTCGCGGCGACCGCGCCACCGTCA 16339
 16340 CGAAGTACCGCGCTGGCGGACGCGCGCAAGGCCACCACTGAAGTCAAGCA 16399
 16400 CGCCTTCACTCCCGCACATGGACCCCATCTCGACGAGTGCAGCGGTGCGCGCGG 16459
 16460 CCTGACCTTCCACGAGCGGTCTATTCCTGCTCTTCAAGTCAACCGTGAAGTGTGAC 16519
 16520 CGGACCGCGACCGGAGCGGCGCGGCGAGCGCGACCGCGAGTACTGGCGCGGCAATGC 16579
 16580 GCGGAGCGCGTGGGTCTGTCTCGGGGTGCGGGGCTGTGCGAGCGCGGGGTGACCA 16639
 16640 GTTGTGAGCTCGGCGCGGACGACCGCTGTCCGCGATGCGCGGACTGCTTCCCGC 16699
 16700 CCGCGGACCGGAGCGCTCGCGCGCGCGCGCATCGCCACATGCGCGCGGCGCGA 16759
 16760 CGAGGTGGCCACGTTCTGAGGTCTGAGCGCGCGCGAGCGTACGTCCGCGCGCGATGTCGA 16819
 16820 CTTACCGCGGCTACGGCGCCACCGCCACGCGCGCTTCCCGCTCCCGACGATCCCTT 16879
 16880 CCAGCGGAGCGCCATTGGCTGCGCTGCGGCTGCGGGGTGGGCGAGCGCGGAGACCGCGA 16939
 16940 ACTTCCGGAATCTCGGAGTCTCGGAGTCCGAGCAGGCGATGAGCGGAGGAGGGGCGG 16999
 17000 CGCGTGGGCGGGCTGAAGGCGGCTTGCAGGCTCTCCGTGAACGACGAGGCGGT 17059
 17060 CCTCTCGGCTGGTCAACAGCAGTGGCGCTGCTGCTCGGGAGCGCTCGGGACGCT 17119
 17120 ACAAGCGCGCGCACTTCAAGCAGTGGGCTTCGACTCGATGGCGCGCGAGCTGAG 17179
 17180 CGAAGCGCTCGGACGAGCGGCGCTGCGGTGCGCGCGACCTACCTTCGACTACCC 17239
 17240 GACCCCTCTGGCGGTGCGCGGACCTGCGCGGAGTCAACGCTACGCGCGCGCGC 17299
 17300 CGGCTCGCGCGCGCGACGGCGCGCTCGCGCGGGGTGACCTCGGGACGAGGAGCCC 17359
 GGTGCGCCATCG 17370

Fig 5